Folding of Repeat Proteins

Designed Ankyrin Repeat and Cysteine-rich Repeat Protein from *Helicobacter pylori*

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Sathya Devi Venkataramani

aus Indien

Promotionskomitee

Prof. Dr. Hans Rudolf Bosshard (Vorsitz)

PD Dr. Ilian Jelesarov

Prof. Dr. Markus Grütter

Zürich 2005

To my beloved parents Mrs. V. Muthu sankari and
Mr. V. Venkataramani

CONTENTS IN BRIEF

Table of Contents

Summary

Repeat proteins are ubiquitous and are involved in many biological reactions. They have non-globular structures acting as scaffolds for protein–protein interactions. A repeat protein is composed of tandem repeats of a basic structural motif of 20–40 amino acids. It is mainly stabilized by hydrophobic interactions within the single repeats and between adjacent repeats. Long-range tertiary contacts, which are the hallmark of globular proteins, are very rare in repeat proteins. The modular structure poses exigent questions about the folding of repeat proteins. Does the modular nature affect the folding mechanism? Do the repeating motifs fold independently, leading to tangible folding intermediates? To explore the folding properties of repeat proteins, we selected an ankyrin repeat protein and a cysteine-rich repeat protein from *Helicobacter pylori*. The ankyrin motif has 33 residues arranged in a β-turn followed by two α-helices connected by a loop. The cysteine-rich motif has 36 residues arranged in two α-helices connected by a disulfide bridge.

The ankyrin repeat protein was selected from a consensus library of ankyrins with only a single repeat flanked by N- and C-terminal capping repeats. I found that this simplest possible ankyrin repeat folds in a two-state manner at 5 °C, similar to many small globular proteins. However, folding is more complex at 25 °C and above where equilibrium intermediates are observed. The results are described in Chapter 2 and are based on thermal melting and urea-induced unfolding studies (both kinetic and equilibrium) and on differential scanning calorimetry (DSC).

A family of cysteine-rich repeat proteins (Hcp) was found in *Helicobacter pylori*, a major risk factor for several gastric diseases. A single disulfide bridge within each repeat is a unique feature of this class of repeat proteins. I have studied the folding of the protein HcpB with four repeats and four consecutive disulfide bridges. The results are described in Chapter 3. The disulfide bridges of HcpB are arranged one after the other along the amino acid sequence of HcpB, an arrangement well suited to demonstrate independent folding of the four repeats within HcpB. From studies on disulfide-containing globular proteins we know that wrong, that is, non-native disulfide bridges can form during folding and have to be re-arranged in order to reach the native disulfide pattern. In the case of a repeat protein, wrong disulfide bonds should not appear if the repeats fold in an independent way before disulfide formation. I have followed *in vivo* oxidative folding of HcpB in the periplasm of *E. coli*, which has an enzyme system for disulfide formation (DsbA/DsbB) and another one for re-arranging wrong disulfide bonds (DsbC/DsbD). I find that the folding of HcpB needs only DsbA/DsbB. Hence, no wrong disulfide bonds are formed during

the folding of HcpB in the periplasm of *E. coli*. Obviously, a repeat-like folding pattern develops before disulfide formation, circumventing wrong disulfide formation.

Folding of HcpB in the test tube was followed in the presence of a glutathione redox buffer. Here again, no wrong disulfide formation is seen. The result is based on mass-spectrometric analysis of folding intermediates with one, two and three disulfide bonds. Folding proceeds through one major folding intermediate with three native disulfide bonds in repeats 2, 3 and 4 and a reduced pair of cysteines in repeat 1. Establishment of the disulfide bond in repeat 1 limits the overall folding rate of HcpB and contributes almost half of the thermodynamic stability of HcpB. Taken together the results described in Chapters 2 and 3 provide a first glimpse on how unfolded polypeptide chains fold into the modular structure of repeat proteins.

In Chapter 4, I describe experiments on the effect of urea and guanidinium chloride (GdmCl) on the stability of a leucine zipper (coiled coil). The free energy of proteins can be deduced from the effect of chemical denaturants on protein stability. Data analysis follows the linear extrapolation model (LEM). Occasionally, LEM provides different values of stability depending on the nature of the chemical denatured used. I have tried to solve this problem on the example of leucine zipper AB_{SS} designed in our laboratory. This two-stranded leucine zipper features several electrostatic interactions between charged amino acid side chains and provides a good model to study the effect of urea and GdmCl, the two most widely used denaturants, on protein electrostatics. Previously, we had shown that GdmCl unfolding leads to a higher free energy of unfolding of AB_{SS} at pH 2 than unfolding with urea. To clarify this puzzling observation, the kinetics of folding of AB_{SS} was measured in the presence of urea and GdmCl at pH 2 and pH 7. The apparent higher stability deduced from GdmCl unfolding at low pH can be plausibly explained by a thousand-fold slower rate of unfolding, indicating strong charge screening of unfavorable charge-charge repulsions at pH 2 by GdmCl but not by urea.

Zusammenfassung

Repeat Proteine sind weit verbreitet und an vielen biologischen Reaktionen beteiligt. Sie besitzen typischerweise eine nicht-globuläre Struktur und wirken als Gerüste für Protein-Protein Wechselwirkungen. Ein Repeat Protein besteht aus aneinander gereihten, sich repetierenden Motiven („Repeats“), die je 20–40 Aminosäurereste enthalten. Die Moleküle werden vorwiegend durch hydrophobe Kräfte innerhalb der einzelnen Repeats und zwischen benachbarten Repeats stabilisiert. Kontakte zwischen weit auseinander liegenden Sequenzabschnitten, wie sie für globuläre Proteine typisch sind, kommen in Repeat Proteinen selten vor. Aus der modularen Struktur ergeben sich Fragen zum Faltungsmechanismus. Wie beeinflusst die modulare Struktur den Faltungsmechanismus? Falten sich die Repeats unabhängig voneinander, was zu fassbaren Zwischenprodukten führen dürfte? Um die Faltung von Repeat Proteinen zu erforschen, haben wir ein Ankyrin Repeat Protein und ein Cystein-reiches Repeat Protein aus *Helicobacter pylori* untersucht. Das Ankyrinmotiv besteht aus 33 Resten, die in einer β-Schleife gefolgt von zwei über einen Loop verbundene α-Helices angeordnet sind. Das Motiv des Cystein-reichen Repeats hat 36 Reste und besteht aus zwei Disulfid-verknüpften α-Helices.

Das Ankyrin Repeat Protein wurde aus einer Konsens-Bibliothek ausgewählt, die Ankyrine mit einem einzigen Repeat zwischen N- und C-terminalen „Repeat-Kappen“ (capping repeats) enthält. Dieses einfachste Ankyrin faltet bei 5 °C nach einem einfachen 2-Zustandsmechanismus, ganz ähnlich wie viele kleine, globuläre Proteine. Allerdings ist der Faltungsmechanismus bei 25 °C und darüber komplizierter, denn bei höherer Temperatur werden Faltungsintermediate beobachtet. Die Resultate sind im Kapitel 2 beschrieben. Sie basieren auf thermischer und chemischer Denaturierung (kinetisch und Gleichgewichtsmessungen) und auf differentieller Scanning-Kalorimetrie (DSC).

Eine Familie von Cystein-reichen Repeat Proteinen (Hcp) wurde in *Helicobacter pylori*, einem Risikofaktor für Erkrankungen des Magens, gefunden. Eine einzelne Disulfidbrücke in jedem Repeat ist ein besonderes Merkmal dieser Proteinklasse. Ich habe die Faltung von HcpB, einem Protein mit vier Repeats und vier Disulfidbrücken, untersucht. Die Resultate sind im Kapitel 3 beschrieben. Die Disulfidbrücken von HcpB erscheinen der Reihe nach in der Aminosäurensequenz. Aus der Untersuchung von globulären, disulfidhaltigen Proteinen ist bekannt, dass im Laufe der Faltung falsche, d. h. nicht-native Disulfidbrücken entstehen, die zur Erlangung des nativen Disulfidmusters nachträglich umgeordnet werden müssen. Bei der Faltung von Repeat Proteien, die in jedem Repeat eine Disulfidbrücke aufweisen, sollten keine nicht-

nativen Disulfidbrücken entstehen, vorausgesetzt die Repeats entstehen unabhängig voneinander und vor der Disulfidbildung. Ich habe die oxidative Faltung von HcpB *in vivo* im Periplasma von *E. coli* untersucht. Das Bakterium besitzt ein Enzymssystem zur Bildung von Disulfidbindungen (DsbA/DsbB) und eines für das Umordnen falscher Disulfidbindungen (DsbC/DsbD). Die Faltung von HcpB benötigt nur das System DsbA/DsbB. Demzufolge entstehen während der Faltung von HcpB im Periplasma keine falschen Disulfidbindungen. Offensichtlich entsteht ein repetitives Strukturmuster schon vor den Disulfidbindungen.

Die Faltung von HcpB im Reagensglas untersuchte ich in Gegenwart von Glutathion-Redoxpuffer. Auch hier entstehen keine falschen Disulfidbindungen. Das Resultat wurde mit Hilfe der massenspektrometrischen Analyse von Faltungsintermediaten mit einer, zwei und drei Disulfidbindungen erhoben. Die Faltung erfolgt über ein Hauptzwischenprodukt, das in den Repeats 2, 3 und 4 native Disulfidbindungen und in Repeat 1 zwei reduzierte Cysteine enthält. Die Bildung der Disulfidbindung im 1. Repeat bestimmt die Faltungsgeschwindigkeit des gesamten Proteins und trägt beinahe die Hälfte der Stabilität von HcpB bei. Die Resultate in den Kapiteln 2 und 3 geben einen ersten Einblick in die Art und Weise, wie sich ungefaltete Polypeptidketten zu modularen Strukturen in Repeat Proteinen auffalten können.

Im Kapitel 4 beschreibe ich Experimente zum Einfluss von Harnstoff und Guanidiniumchlorid (GdmCl) auf die Stabilität eines Leuzin-Zippers (Coiled Coil). Die freie Enthalpie von Proteinen lässt sich aus dem Einfluss von chemischen Denaturantien auf die Proteinstabilität herleiten. Die Datenanalyse geschieht nach dem Modell der linearen Extrapolation (LEM). Allerdings kommt es vor, dass die linear Extrapolation für verschiedene Denaturantien verschiedene Stabilitätswerte ergibt. Ich habe versucht, dieses Problem für den Leuzin-Zipper AB_{SS}, einem Proteindesign aus unserer Arbeitsgruppe, zu lösen. Der zweisträngige Leuzin-Zipper zeigt mehrere elektrostatische Wechselwirkungen zwischen geladenen Aminosäureseitenketten. Er ist deshalb ein gutes Modell, um den Einfluss der beiden häufig verwendeten Denaturantien Harnstoff und GdmCl auf die Proteinelektrostatik zu studieren. Wir hatten zuvor gezeigt, dass bei pH 2 die GdmCl-Denaturierung eine grössere Stabilität als die Harnstoff-Denaturierung ergibt. Um diesen rätselhaften Befund zu erklären, wurde die Faltungskinetik von AB_{SS} in Gegenwart von Harnstoff und GdmCl bei pH 2 und pH 7 gemessen. Die aus der GdmCl-Denaturierung abgeleitete höhere Stabilität kann mit der tausendfach langsameren Denaturierungsgeschwindigkeit in Gegenwart von GdmCl bei pH 2 plausibel erklärt werden. GdmCl, nicht aber Harnstoff, dämpft energetisch ungünstige Ladungsabstossungen bei pH 2 und erhöht dadurch die Stabilität von AB_{SS} im sauren pH-Bereich.

1 Protein Folding and Repeat Proteins

1.1 Protein folding

Understanding protein folding means understanding the molecular processes by which the one-dimensional sequence information of a polypeptide chain is transformed into the three-dimensional structure of a biologically active protein. This is a topic of tremendous biological importance and intellectual interest [1-3]. Significant advances have been made over the years, ranging from Anfinsen's original work on ribonuclease to the "new view" of protein folding describing the process as a summation of multiple microscopic pathways on an energy landscape. Folding can be studied *in vitro* and important advances have evolved from experimentation under artificial conditions outside of the living cell. Sections 1.1.1 to 1.1.4 give a brief historical summary. In nature, most of protein folding is an intracellular process, which may be very different from protein folding in the test tube. As a result, *in vivo* studies have matured in parallel to the *in vitro* folding studies (section 1.2).

Understanding protein folding has practical applications in medicine and biotechnology. Expression of recombinant proteins is often hampered by low yields, misfolding and aggregation. Here, a deeper understanding of protein folding could be beneficial. Also, misfolding and aggregation of proteins plays a role in the development of several pathologies of which BSE and Alzheimer's disease are prominent examples [4].

1.1.1 The beginnings

Around 1930, at a time when proteins were mainly envisioned as gelatinous matter without defined molecular structure, the first studies on protein denaturation and on the recovery of the native state appeared in the literature [5]. Several papers were published in the 1930's demonstrating reversibility of protein unfolding. This has led to the concept that protein folding is a spontaneous process. In the 1950's, new thermodynamic insight underlined the importance of non-covalent interactions in proteins. An important concept arose in 1959 when Kauzmann suggested that the hydrophobic effect is the driving force directing the folding process [6]. Starting in the 1960's, the determination of the 3D structure of proteins by X-ray diffraction provided a new basis for structure analysis and for studying the folding process [7].

The Anfinsen postulate – In 1961 Anfinsen and his co-workers published a seminal paper on the spontaneous oxidative refolding of reduced and unfolded ribonuclease into active enzyme. This marked the beginning of modern research into protein folding. Anfinsen's conclusion was succinct: "*All the information necessary to achieve the native conformation of a protein in a given environment is contained in its amino acid sequence*" [8]. The corollary of this postulate is the thermodynamic control of protein folding, meaning that the native conformation is at the minimum of the Gibbs free energy.

The Levinthal paradox – Cyrus Levinthal invoked an influential argument in 1968 that came to be known as Levinthal's paradox, stating that due to the vast number of conformations a polypeptide can adopt, it will not fold in a reasonable amount of time if it has to randomly search the entire conformational space [9]. To resolve the paradox it was assumed that proteins travel only a limited subset of conformations or that folding proceeds along a unique and well-defined folding pathway, akin to a chemical reaction pathway between small molecules. Levinthal and later Wetlaufer [10] have regarded protein folding to be under kinetic rather than thermodynamic control. Kinetic control means that folding is rapid (seconds) because it is pathway dependent, that is, the final structure may depend on the denatured state from where folding is initiated. Hence, the folded protein may occupy some local energy minimum instead of the global minimum. In contrast, thermodynamic control implies that a folding protein is at its global energy minimum and that, therefore, folding has to be pathway-independent and the native structure depends on final conditions, not on the conditions from where folding starts. Time is not relevant in thermodynamically controlled folding and it may take long before the global energy minimum is reached through an extensive conformational search.

Levinthal's paradox has been likened to a ball rolling on an entirely 'flat playing field' or 'golf-course' before dropping to the bottom of a single, small hole representing the global energy minimum of the native state [11] (Fig. 1.1).

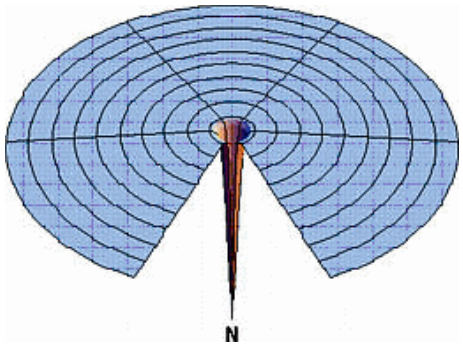


Figure 1.1. The Levinthal 'golf-course' landscape. N is the native conformation. The peptide searches for N randomly, as if a blind golf player would be playing on a totally flat golf-course. Figure adapted from [11].

1.1.2 The classical view of protein folding

Levinthal's paradox led to a search for folding pathways. Folding along a predetermined pathway is considered the classical view of protein folding. Folding is assumed to start from the denatured state (D) and to proceed through a series of well-defined steps to the native state (N). Fig. 1.2 depicts the folding pathway as a downhill canyon with the denatured state D at its start. The canyon may contain valleys and hills representing intermediate and transition states between D and N. In molecular terms, the pathway is a well-defined sequence of events such as a series of changes in dihedral angles. Most important, the conformational search occurs only along the canyon, which limits the search drastically and makes the folding reaction fast.

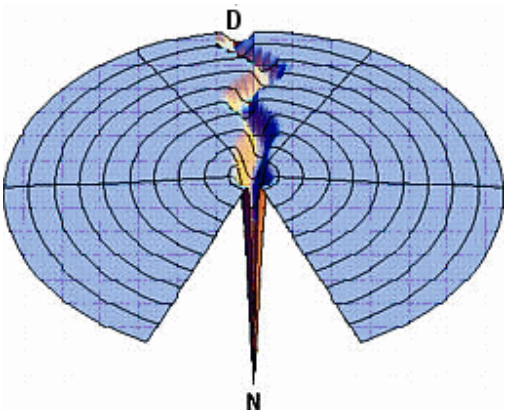


Figure 1.2. The 'pathway' solution to the random search problem of Fig. 1.1. The denatured state is represented as D and the native state as N. Folding of the polypeptide chain occurs along a predetermined pathway depicted as a canyon in the folding landscape. Figure adapted from [11].

Inherent in the view of a folding pathway is the concept of folding intermediates. It was thought that to learn how nature sieves so quickly through the conformational haystack one has to

identify intermediate states along the folding pathway. The pioneering papers of 1971 by Ikai and Tanford [12] and Tsong, Baldwin and Elson [13] mark the beginning of the search for folding intermediates. Major impediments in characterizing folding intermediates are the high cooperativity of protein folding and its high speed, which often is in the time range of milliseconds. Nevertheless, much effort has been devoted to the characterization of transient folding intermediates and a plethora of techniques and experimental designs has been invoked in their search. One of the earliest method was kinetic trapping of intermediates during the refolding of disulfide bridged proteins, as developed by Wetlaufer [14, 15, 16]. Another early method was pulsed chemical labelling introduced by Ghélis [17]. Stopped-flow mixing coupled to circular dichroism or fluorescence spectroscopy, and NMR spectroscopy combined with pulsed hydrogen-deuterium exchange of transient species provide information at the single amino acid residue level [18-20]. Transient folding intermediates can be accumulated by working with protein fragments that are unable to complete the folding process. More generally, protein engineering can be employed to stabilize intermediates or to probe particular regions of a folding protein. Outstanding among the latter methods is the analysis of Φ-values developed by Fersht and co-workers [21-23].

1.1.3 The new view of protein folding

While the classical view of protein folding relies on phenomenological models and regards folding intermediates in a structural way, the new view emphasizes the ensemble nature of protein conformations. Folding is no longer thought to start from the denatured state since D is not regarded as a defined starting state like the starting conformation of a small molecule from which a chemical reaction initiates. Rather, the denatured state comprises of a very large ensemble of states. As a consequence, folding is viewed as a large ensemble of parallel pathways involving diffusion-like processes. The folding protein “funnels” to the global minimum state by multiple routes in conformational space, which is considered as an energy landscape (Fig. 1.3). The Levinthal paradox disappears in the new view of protein folding since the energy landscape is not a “flat golf course” but a “downhill slope”.

An idealized smooth funnel based on an early mean-field lattice model [24] is represented in Fig. 1.3A. The figure shows how funnels resolve Levinthal’s paradox. An analogy can be drawn between the denatured state and an ensemble of skiers distributed at the top of a mountainside

(the rim of the funnel). When folding conditions are initiated, each skier glides downhill on his own individual trajectory. Skiers reach the global minimum (satisfying Anfinsen's postulate) by many different routes rather than by a single route or folding pathway. Yet they do so in a directed and rapid way (easing Levinthal's concern).

Fig. 1.3B shows a rough energy landscape with kinetic traps arising from energy barriers; in this situation slow and multi-state folding is observed. When local barriers are high, protein molecules can be trapped and, depending on the nature of the states, may aggregate. Alternatively, some unfolded peptides may fall into a trap (local kinetic minimum) that is not on a pathway leading to the native state and an off-pathway intermediate is formed (Fig. 1.4). Peptide B in Fig. 1.4 can exit the local energy minimum and get back to the folding track.

In the words of Dobson and Karplus [25] the new view of protein folding *"... provides a simple way of understanding why the Levinthal paradox is not a real problem*". Indeed, the energy landscape and the folding funnel metaphor are a conceptual framework for understanding the different scenarios of protein folding as well as misfolding and aggregation.

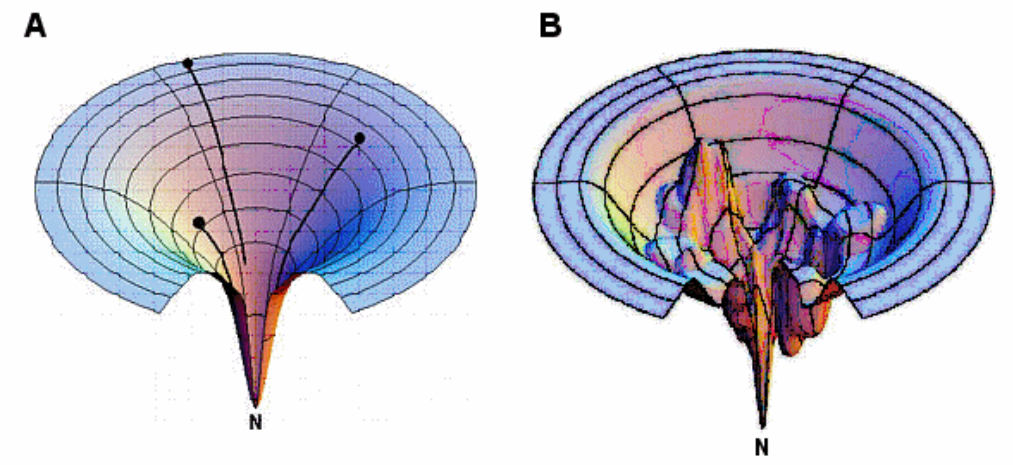


Figure 1.3. Smooth and rough funnels. (A) A smooth funnel represents a two-state folding mechanism from any unfolded state (any position on the funnel rim) to the folded state N. There is no folding intermediate and folding is very fast. (B) A rugged energy landscape represents kinetic traps, energy barriers and some narrow throughway paths. Folding is multi-state and proceeds through intermediates. Figure adapted from [11].

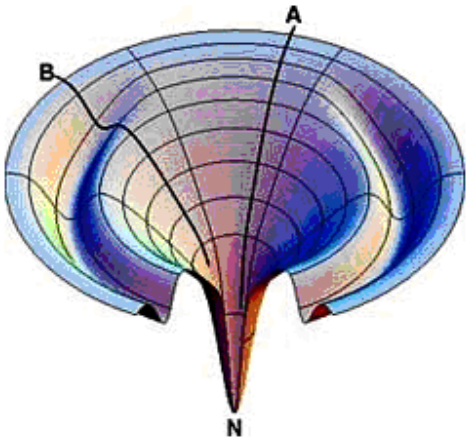


Figure 1.4. A folding landscape to illustrate off-pathway intermediates. Protein A folds by a fast throughway process. Protein B gets trapped in a local energy minimum from which it may escape to reach the same global energy minimum as protein A, yet more slowly. Figure adapted from [11].

1.1.4 Folding models

Different folding models have attained prominence. They originated from the classical pathway view of protein folding. Fig.1.5 shows four long-standing and well-established models. Having grown out from theoretical considerations [26-28], molecular simulations [29-32] and experimental observations [33-35], the models try to explain the high speed of protein folding. Although it is now clear that there is not a single sequential folding route, as is implicit in some of these folding models, many features of the models are still valid in the context of the new view of protein folding by multiple folding routes in conformational space.

The nucleation-condensation model – The model of nucleation-propagation, which applies to the helix-coil transitions [36, 37], involves a nucleation step followed by rapid propagation to the folded state. The nucleation process is considered rate-limiting. The nucleation-condensation model proposed by Fersht [38] is a corollary of the older nucleation-propagation model; weak local folding nuclei are formed and are stabilized by long range interactions. The nucleation-condensation model is experimentally well supported for several small proteins including chymotrypsin inhibitor II [39, 40] and barstar [41].

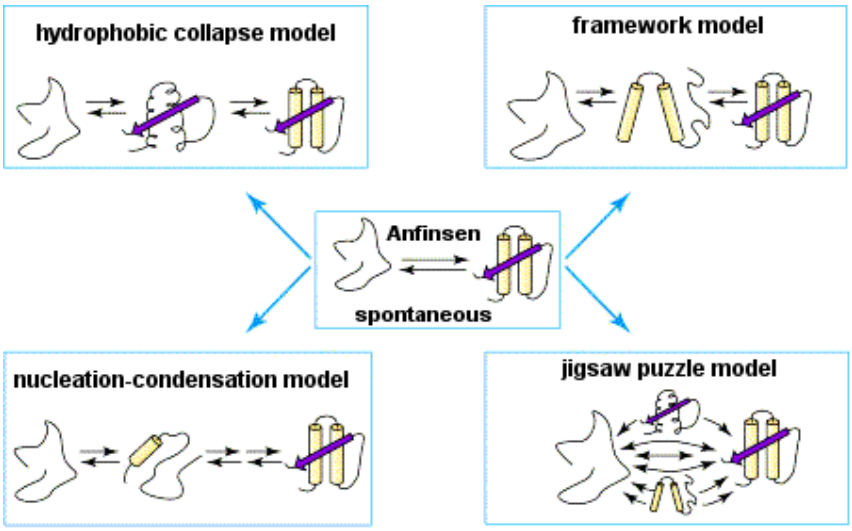


Figure 1.5. Four prominent folding models. Each model attempts to explain why folding is rapid or why Levinthal's paradox is mistaken. See the text for details. Figure adapted from [42].

The framework model (diffusion-collision model) – According to this model, local elements of secondary structure form first (usually within a few μs) and these then dock to form the native tertiary structure of the protein, possibly by a diffusion–collision mechanism. The model emphasizes the role of short range interactions in directing protein folding [43]. A sequential and hierarchical model of protein folding has long been generally accepted [33, 44].

The hydrophobic collapse model (molten globule) – In this model, the protein buries its hydrophobic side chains from solvent water early during folding, forming a collapsed intermediate also known as a molten globule. The native state develops by a search within the conformationally restricted area of the molten globule. This model considers that the first event of protein folding consists of a collapse that occurs via long range interactions, before or at the same time as the formation of secondary structures [24]. The model was inspired by Kauzmann's hydrophobic effect [6].

The jigsaw puzzle model – The model was introduced by Harrison and Durbin in 1985 [45]. These authors proposed the existence of multiple protein folding pathways, as in the assembly of a jigsaw puzzle in which multiple routes exist to reach a unique solution. According to the model,

the identification of intermediates attempts a kinetic rather than a structural description, each intermediate consisting of heterogeneous species in rapid equilibrium with each other.

1.2 Protein folding in the cell

The main rules of protein folding have been deduced from *in vitro* and *in silico* studies. Although it has been accepted since Anfinsen's work that *in vitro* refolding is a valid model to understand the folding of a nascent polypeptide chain *in vivo*, there are important differences between the cell and the test tube. The interior of a cell is highly crowded with macromolecules. The total protein concentration inside a cell is 200 – 300 mg mL^{-1}, that of RNA is 300 – 400 mg mL^{-1}, and polysaccharides also contribute to molecular crowding [46]. Moreover, inside a cell the newly synthesized polypeptides enter the cytosol vectorially because the N terminus is synthesized before the C terminus. This means the N terminus of a nascent polypeptide is available for folding before the rest of the polypeptide, yet folding units or domains cannot occur until the chain has reached a certain length [47, 48]. So the question remains: Do the same mechanisms account for protein folding *in vitro* and *in vivo*?

1.2.1 The role of molecular chaperones

Protein folding in the cell is strongly hampered by molecular crowding. As a result, partly folded proteins with hydrophobic surface still exposed to the aqueous surrounding may accumulate, which may lead to misfolding or aggregation. The efficiency and speed of cellular protein synthesis is reduced. Chaperons counteract misfolding and aggregation [49]. These helper molecules have been intensively studied and significant progress has been made in the understanding of how chaperones assist the folding process [50, 51]. The first molecular chaperone identified was nucleoplasmin involved in mediating the assembly of nucleosomes [52]. Over 20 chaperone families have since been identified. Some of the best known are the heat-shock proteins Hsp 70 (DnaK in *E.coli*) and Hsp 40 (DnaJ in *E.coli*). They show little or no specificity for the proteins they assist in folding [53]. They bind transiently to small hydrophobic regions of nascent polypeptide chains, preventing aggregation and premature folding. Release of the bound protein is often, but not always, regulated in a complex ATP-dependent way.

The large chaperonins such as the system GroEL-GroES in prokaryotes or the TCP-1 Ring Complex (TRiC) in eukaryotes, sequester the unfolded or partially folded polypeptide chain in a

central cage. Electron microscopy, high resolution crystallography [54, 55], kinetic studies and dynamic fluorescence [56-58] have provided information on the GroEL-GroES machinery. In essence, chaperonins smooth the energy landscape for folding by encapsulating the polypeptide chain in a folding-friendly environment within the GroEL-GroES macromolecular complex. Such behavior has been confirmed in simulations of chaperonin-assisted folding, in accord with experimental results [59, 60].

1.2.2 Misfolding, aggregation and pathological consequences

Protein aggregation is a widespread phenomenon, often arising from early folding intermediates through kinetic competition between proper folding and misfolding. The generally accepted hypothesis for protein aggregation is that either in the folding process or as a result of fluctuations partially folded intermediates become populated. In the intermediates some hydrophobic regions are exposed, thus permitting aggregation. The morphology of aggregates can be ordered and leading to amyloid fibrils, or amorphous and resulting in inclusion bodies [61].

Chaperones alleviate the problem of protein misfolding or aggregation. Still, situations exist where chaperones fail, for example under severe stress or when mutant proteins exhibit aberrant folding or stability properties. Occasionally, protein misfolding and aggregation can lead to diseases, some of them fatal. The formation of amyloid fibrils plays a key role in the origin of several pathologies such as spongiform encephalopathies. There are severe neurodegenerative diseases associated with an abnormal prion proteins: Kuru, Creutzfeld-Jacob disease (CJD) and fatal familial insomnia (FFI) in humans, scrapie in sheep, and bovine spongiform encephalopathy (BSE) in cattle. Alzheimer's disease is also characterized by the presence of amyloid febrile deposits in brain tissue. Recent observations have shown that several proteins unrelated to amyloid diseases, such as ordinary lysozyme or myoglobin, are able to aggregate *in vitro* into amyloid fibrils that are indistinguishable from those found in pathologic conditions [62-64]. Furthermore, species formed early in the aggregation of non-disease associated proteins can be cytotoxic [65]. Therefore, the control of misfolding and aggregation is of fundamental importance for cell viability.

In conclusion, it is clear that over the past 25 years our concepts of folding have advanced dramatically. As is typical of good research, new insights elicit new questions. Hence, there are

plentiful opportunities to dig deeper into the folding problem, both *in vitro* and *in vivo*. Advances of the future will rely on the synergy between experimental and computational approaches.

1.3 Repeat proteins

A large number of protein classes are built on the modularity principle from homologous structural blocks [69, 70]. Repeating stretches of similar sequence are found in at least 14% of all proteins [69]. The repeats vary from a few amino acids, for example, the polyglutamine tracts of the Huntington disease gene product huntingtin, to large repetitions of multiple domains as found in the cytoskeletal protein titin. Repeats arise via intragenic duplication and recombination. Multiplication of preexisting genetic material enables an organism to expand its repertoire of cellular functions, such as protein transport, protein-complex assembly and protein regulation. Indeed, the most common function of repeats is to bind other proteins.

Repeat proteins are found in all phyla but are most common in eukaryotes, particularly in metazoans [69]. This may have to do with the increasing complexity of cellular functions of higher organisms. One attractive hypothesis is that repeat proteins evolve more quickly than non-repeat proteins. Eukaryotes possess a sophisticated protein synthesis machinery to handle the multi-domain, non-globular folds of repeat-rich proteins. Eukaryotes may use this as an advantage over prokaryotes to gain from the benefits offered by repeats: modular construction of new proteins and introduction of rapidly evolving protein sequences that allow faster adaptation to new environments [69].

1.3.1 Structural organization of repeat proteins

Repeat proteins consist of a number of structurally identical motifs usually arranged in tandem and stacked together to form elongated or supercoiled shapes [71]. Repeating modules are typically 20-40 residues long and contain secondary structure elements that fold in a variety of topologies [72]. The linear assemblage of complementary repeats results in a relatively simple and robust scaffold, which is maintained by the regular repetition of hydrophobic contacts and hydrogen bonds. The architecture of stacked elements implies that stabilizing contacts are either within one repeat or between adjacent repeats. There are hardly any contacts between residues of non-adjacent repeats.

1.3.2 Major classification of repeat proteins

Repeat proteins have been classified into six families according to secondary and tertiary structure elements [73].

secondary structure	repeat families
all α structure	armadillo/HEAT
	tetratricopeptide repeats
all β structure	β-propellers
	β-trefoils
mixed α/β structure	leucine-rich repeats
	ankyrin repeats

Repeats containing only α-structures – The 42 amino acid **armadillo repeat** (arm motif) [74] was identified in the product of the *Drosophila melanogaster* segment polarity gene [75]. It has since been found in several eukaryotic proteins including β-catenin, importin α, the tumour suppressor adenomatous polyposis coli (APC), etc. The arm motif consists of three α-helices associated to a superhelical structure. The first helix is short (about 8 residues) and lies perpendicular to the other two longer α-helices that pack against one another. The function of tandem arm repeats is to bind proteins, which is achieved by a long groove extending along the superhelix axis created by the repetitive arm motifs. Fig. 1.6 shows the nuclear import receptor importin α containing 10 arm repeats [76].

The **HEAT** motif (Fig. 1.7) is a repetitive sequence that was first observed to be common to the **H**untingtin protein, **e**longation factor 3, PR65/**A** subunit of protein phosphatase A, and the target of rapamycine (**T**OR) [77]. HEAT motifs vary in length between 37 and 43 residues and occur in tandem arrays of between 3 and 22 motifs. Occasionally blocks of tandem repeats are dispersed throughout the sequence. HEAT motifs are also present in importins β1 and β2, in proteins related to the clathrin-associated adaptor complex and in many other proteins related to chromosome dynamics [78]. HEAT motifs are involved in protein-protein interactions. HEAT repeats have two anti-parallel α-helices. The first HEAT helix has a kink (of variable extent) that

makes it equivalent to both the first and second helices of armadillo repeats. The parallel stacking of repeat units produces a solenoid.

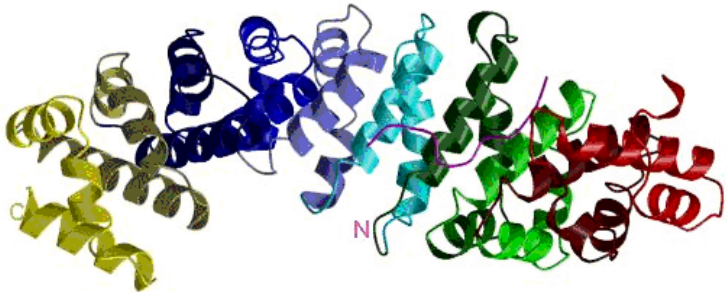


Figure 1.6. Armadillo repeats of importin alpha (PDB code: 1IAL). Different arm repeats are colored differently; the N-terminal auto inhibitory segment is in magenta [76].

Figure 1.7. Ribbon diagram of the transport factor karyopherin-β2 with 18 HEAT repeats (PDB code: 1QBK) [79]. Repeats are differently coloured and labelled HR1–HR18. The N-terminal capping helices NαA and NαB) are red and the C-terminal capping helices CαA and CαB are pink. Loop L7 is dark green; it connects the helices of HR7 with the short α-helical regions α1, α2 and α3.

The **tetratricopeptide repeat (TPR)** was identified in 1990, the name denoting the 34 amino acids of the basic repeat [80, 81]. The 34 amino acids are arranged in two α-helices packed together in a knobs-in-holes manner. The TPR is likely to be an ancient repeat since it is found in

eukarya, bacteria, and archaea [82]. Multiple TPRs form a right-handed super helix and show multiple modes of ligand binding in a large groove [83]. Fig. 1.8 shows the crystal structure of a designed TPR motif [84]. TPRs come in many different flavours that form distinct sequence subfamilies. They are present in kinesin light chains, SNAP secretory proteins, clathrin heavy chains and bacterial aspartylphosphate phosphatases. Malfunctioning of a TPR-protein is linked to human diseases [85, 86].

Figure 1.8. Crystal structure of designed TPR repeat protein CTPR3 consisting of three repeats (PDB entry: 1NA0). Repeat 1 is yellow, repeat 2 is red, repeat 3 is blue, and the capping helix is green [84].

Repeats containing only β-structures – The WD40 repeat, also known as a **β-propeller**, is the most common repeat among the known human proteins [87]. The motif contains approximately 40 residues and includes well-conserved tryptophan and aspartic acid residues. The crystal structure of an assembly of seven WD40 repeats (Fig. 1.9) reveals that each repeat represents a four-stranded antiparallel β-sheet (a "blade") arranged radially in a "propeller" arrangement about a central axis. Seven-blade β-propellers are found in methylamine dehydrogenase (PQQ repeats), regulator of chromosome condensation 1 (RCC1 repeats) and galactose oxidase (Kelch repeats). Neuraminidase features a six–blade propeller. The propeller structure is closed through interactions between the N- and C-terminal repeats. Again, a common function of β-propellers is to bind ligands and other proteins through a ligand binding "supersite" formed by the N-termini of interior β-strands [88].

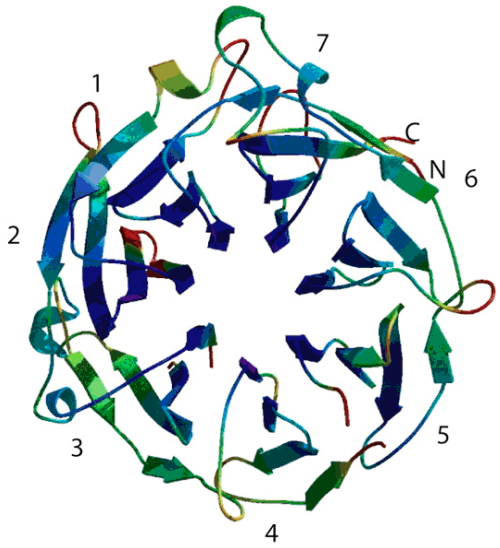


Figure 1.9. Structure of Ski8p, a protein regulating mRNA degradation with a seven-blade β-propeller [89]. Each propeller consists of four anti-parallel β-strands.

β-trefoils are another type of all β-sheet "closed" structures. The β-trefoil has six two-stranded β-hairpins, three of which form a barrel structure while the remaining three form a triangular cap on the barrel [90]. The fold occurs in fibroblast growth factor (FGF), interleukin-1, Kunitz soybean trypsin inhibitor and ricin-like toxin. Four β-trefoils are found in actin-binding fascin. By contrast to the β-propellers, the β-trefoils do not appear to possess a "supersite" for ligand binding since they bind the ligands in different locations [88].

Repeats containing α/β structures – Leucine-rich repeats (LLRs) were first identified in the leucine-rich α2-glycoprotein [91]. LRRs are 20–29 residues long and contain a conserved 11-residue segment and consist of a β strand and an α helix connected by loops [92]. The repeats are arranged so that all the strands and helices are parallel to a common axis, resulting in a nonglobular, horseshoe-shaped molecule with a curved parallel β sheet lining the inner circumference of the horseshoe and the helices flanking the outer circumference, as shown in Fig. 1.10.

LRR proteins participate in hormone–receptor interactions, enzyme inhibition, cell adhesion, cellular trafficking, etc. They are involved in early mammalian development [93], neural development [94], cell polarization [95], regulation of gene expression [96] and apoptosis signaling [97].

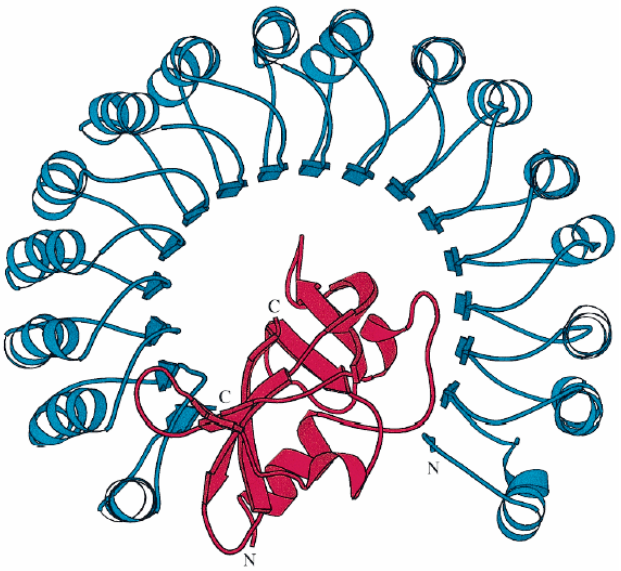


Figure 1.10. Crystal structure of ribonuclease inhibitor RI (PDB code 2BNH). LRRs are shown in blue and an additional subunit is in red.

Ankyrin repeats (AR) take their name from the human erythrocyte protein ankyrin where they were detected first [98]. AR proteins occur in virtually all species, but the majority is found in eukaryotes where they are involved in a wide range of cellular tasks ranging from transcriptional regulation to cytoskeleton organization [99]. The AR is a 33 residues L-shaped motif of two anti-parallel α-helices connected by a short loop [100]. The consecutive repeats stack in parallel and are joined by β-hairpins forming the base of the L. The crystal structure of the ankyrin domain of Bcl-3, a unique member of the IκB protein family [101], is shown in Fig. 1.11. AR proteins bind various other proteins that are not restricted to a single class. There is good evidence for complex formation with p53 (a nuclear tumor suppressor), CDK6 (a cell division kinase) and p65 (a transcriptional regulator). Another interaction is between the developmental AR protein Notch and deltex, a cytoplasmic protein [102]. Some mutations of AR proteins are linked to human diseases [103, 104].

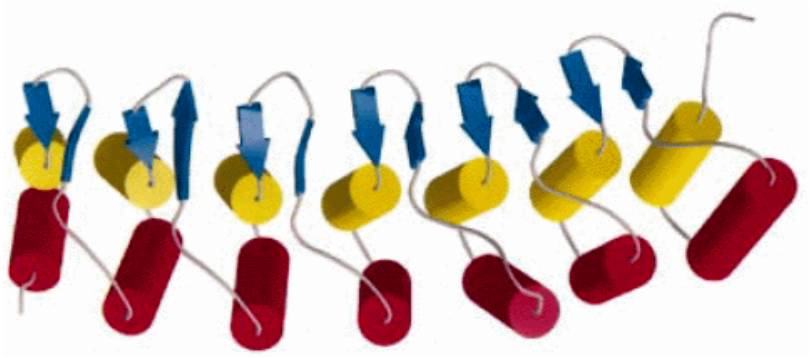


Figure 1.11. Crystal structure of the ankyrin domain of Bcl-3, a member of the IκB family (PDB code: 1K1A). Helices colored red and yellow, β-sheets colored blue.

1.4 Goals of this thesis

Unlike globular proteins, the structure of repeat proteins is dominated by local, short-range interactions. Hence, the folding and stability of repeat proteins may differ from that of similarly sized globular proteins. The goal of this thesis was to look into such potential differences. When I had started this work three years ago, there was very few information available on the kinetics of folding of repeat proteins.

I have chosen two repeat proteins as representative examples for my studies: a small ankyrin with an α/β structure and a small tetratricopeptide-like all-α protein from *Helicobacter pylori* featuring one disulfide bond per repeating motif.

Work on the ankyrin repeat is described in **Chapter 2**. The protein was randomly selected from a designed consensus library constructed by Binz et al. [107]. It is one of the shortest possible ankyrins composed of only a single repeat flanked by N- and C-terminal capping repeats. This simple structure was a good starting point for my investigations. Indeed, under some benign conditions folding can be described as a single-step U →N reaction, much like the folding of a small globular protein. Under other conditions, however, there is evidence for multi-step folding.

Work on the *H. pylori* cysteine-rich repeat protein is described in **Chapter 3**. The protein consists of 4 repeating units and was chosen because of its unique disulfide pattern. Since the disulfide bonds are consecutive, one per repeating unit, we asked if disulfide formation also

proceeds consecutively, or if non-native disulfides appear as transient intermediates to be later reshuffled to the native consecutive order. The answer is unequivocal: Oxidative formation of the disulfide bonds is consecutive both *in vitro* and *in vivo*.

In the final **Chapter 4**, I present my preliminary work on the kinetics of folding of the disulfide-linked heterodimeric leucine zipper AB_{SS} designed in our laboratory to understand the effect of salt bridges on protein stability [109]. It had been shown that leucine zipper AB_{SS} is more stable at acidic pH and that the stability difference between acidic and neutral pH was larger when deduced from GdmCl unfolding than from urea unfolding. Hence, here was a system to look into the longstanding problem of different effects of GdmCl and urea on protein unfolding. The aim was to demonstrate a charge-screening effect of GdmCl as a cause for the different behavior of the two denaturants.

1.5 References

1. Vendruscolo, M., et al., (2003). Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. *Philos Transact A Math Phys Eng Sci* **361,** 1205-22.
2. Dobson, C.M., (2003). Protein folding and misfolding. *Nature* **426,** 884-90.
3. Dobson, C.M., (2004). Principles of protein folding, misfolding and aggregation. *Semin Cell Dev Biol*. **15,** 3-16.
4. Dobson, C.M., (2003). Protein folding and disease: a view from the first Horizon Symposium. *Nat Rev Drug Discov* **2,** 154-60.
5. Wu, H., (1931). Studies on denaturation of proteins.A theory of denaturation. *Adv Protein Chem* **46,** 6-26.
6. Kauzmann, I., (1959). Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **14,** 1-67.
7. Kendrew, J.C., et al., (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181,** 662-6.
8. Anfinsen, C.B., (1973). Principles that govern the folding of protein chains. *Science* **181,** 223-30.
9. Levinthal, C., (1968). Are there pathways for protein folding. *J.Chim.Phys.* **65,** 44-45.
10. Wetlaufer, D.B., (1973). Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci U S A* **70,** 697-701.
11. Dill, K.A. and H.S. Chan, (1997). From Levinthal to pathways to funnels. *Nat Struct Biol*. **4,** 10-9.
12. Ikai, A. and C. Tanford, (1971). Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. *Nature* **230,** 100-2.
13. Tsong, T.Y., R.L. Baldwin, and E.L. Elson, (1971). The sequential unfolding of ribonuclease A: detection of a fast initial phase in the kinetics of unfolding. *Proc Natl Acad Sci U S A* **68,** 2712-5.
14. Wetlaufer, D.B. and S. Ristow, (1973). Acquisition of three-dimensional structure of proteins. *Annu Rev Biochem* **42,** 135-58.
15. Weissman, J.S. and P.S. Kim, (1991). Reexamination of the folding of BPTI: predominance of native intermediates. *Science* **253,** 1386-93.
16. Creighton, T.E., (1978). Experimental studies of protein folding and unfolding. *Prog Biophys Mol Biol*. **33,** 231-97.

17. Ghelis, C., (1980). Transient conformational states in proteins followed by differential labeling. *Biophys J* **32,** 503-14.
18. Baldwin, R.L., (1993). Pulse H/D exchange studies of folding intermediates. *Curr Opin Struct Biol* **3,** 84-91.
19. Dobson, C.M., M. Karplus, and A. Sali, (1998). Protein folding: A perspective from theory and experiment. *Ang. Chem. Int. Ed. Engl.,* **37,** 868-893.
20. Roder, H., G.A. Elove, and S.W. Englander, (1988). Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. *Nature* **335,** 700-4.
21. Matouschek, A. and A.R. Fersht, (1991). Protein engineering in analysis of protein folding pathways and stability. *Methods Enzymol* **202,** 82-112.
22. Garcia, P., et al., (1995). Evidence for residual structures in an unfolded form of yeast phosphoglycerate kinase. *Biochemistry* **34,** 397-404.
23. Ballery, N., et al., (1993). Characterization of an intermediate in the folding pathway of phosphoglycerate kinase: chemical reactivity of genetically introduced cysteinyl residues during the folding process. *Biochemistry* **32,** 708-14.
24. Dill, K.A., (1985). Theory for the folding and stability of globular proteins. *Biochemistry* **24,** 1501-9.
25. Dobson, C.M. and M. Karplus, (1999). The fundamentals of protein folding: bringing together theory and experiment. *Curr Opin Struct Biol* **9,** 92-101.
26. Dill, K.A., et al., (1995). Principles of protein folding--a perspective from simple exact models. *Protein Sci* **4,** 561-602.
27. Karplus, M. and A. Sali, (1995). Theoretical studies of protein folding and unfolding. *Curr Opin Struct Biol* **5,** 58-73.
28. Karplus, M. and D.L. Weaver, (1976). Protein-folding dynamics. *Nature* **260,** 404-6.
29. Kolinski, A. and J. Skolnick, (1994). Monte Carlo simulations of protein folding. I. Lattice model and interaction scheme. *Proteins* **18,** 338-52.
30. Sali, A., E. Shakhnovich, and M. Karplus, (1994). How does a protein fold? *Nature* **369,** 248-51.
31. Taketomi, H., F. Kano, and N. Go, (1988). The effect of amino acid substitution on protein-folding and -unfolding transition studied by computer simulation. *Biopolymers* **27,** 527-59.
32. Levitt, M. and A. Warshel, (1975). Computer simulation of protein folding. *Nature* **253,** 694-8.
33. Kim, P.S. and R.L. Baldwin, (1990). Intermediates in the folding reactions of small proteins. *Annu Rev Biochem* **59,** 631-60.
34. Kim, P.S. and R.L. Baldwin, (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem*. **51,** 459-89.
35. Baldwin, R.L., (1975). Intermediates in protein folding reactions and the mechanism of protein folding. *Annu Rev Biochem* **44,** 453-75.
36. Zimm, B.H. and J.K. Braggs, (1959). Theory of the phase transition between helix and random coil polypeptide chains. *J.Chem.Phys.* **31,** 526-535.
37. Lifson, S. and A. Roig, (1961). On the theory of helix-coil transition in polypeptides. *J.Chem.Phys.* **34,** 1963-1974.
38. Fersht, A.R., (1997). Nucleation mechanisms in protein folding. *Curr Opin Struct Biol* **7,** 3-9.
39. Jackson, S.E. and A.R. Fersht, (1991). Folding of chymotrypsin inhibitor 2. 2. Influence of proline isomerization on the folding kinetics and thermodynamic characterization of the transition state of folding. *Biochemistry* **30,** 10436-43.
40. Jackson, S.E. and A.R. Fersht, (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **30,** 10428-35.
41. Nolting, B., et al., (1997). The folding pathway of a protein at high resolution from microseconds to seconds. *Proc Natl Acad Sci U S A* **94,** 826-30.
42. Radford, S.E., (2000). Protein folding: progress made and promises ahead. *Trends Biochem Sci* **25,** 611-8.

43. Ptitsyn, O.B. and A.A. Rashin, (1973). Stagewise mechanism of protein folding. *Doklady Akademii Nauk SSSR* **213,** 473-475.
44. Jaenicke, R., (1987). Folding and association of proteins. *Prog Biophys Mol Biol* **49,** 117-237.
45. Harrison, S.C. and R. Durbin, (1985). Is there a single pathway for the folding of a polypeptide chain? *Proc Natl Acad Sci U S A* **82,** 4028-30.
46. Ellis, R.J., (2001). Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol* **11,** 114-9.
47. Creighton, T.E., (1990). Protein folding. *Biochem J* **270,** 1-16.
48. Jaenicke, R., (1991). Protein folding: local structures, domains, subunits, and assemblies. *Biochemistry* **30,** 3147-61.
49. Ellis, J., (1987). Proteins as molecular chaperones. *Nature* **328,** 378-9.
50. McClellan, A.J. and J. Frydman, (2001). Molecular chaperones and the art of recognizing a lost cause. *Nat Cell Biol* **3,** E51-3.
51. Frydman, J., (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu Rev Biochem* **70,** 603-47.
52. Laskey, R.A., et al., (1978). Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature* **275,** 416-20.
53. Mayer, M.P. and B. Bukau, (2005). Hsp70 chaperones: cellular functions and molecular mechanism. *Cell Mol Life Sci* **62,** 670-84.
54. Xu, Z., A.L. Horwich, and P.B. Sigler, (1997). The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. *Nature* **388,** 741-50.
55. Braig, K., et al., (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 A. *Nature* **371,** 578-86.
56. Weissman, J.S., et al., (1996). Characterization of the active intermediate of a GroEL-GroES-mediated protein folding reaction. *Cell* **84,** 481-90.
57. Ranson, N.A., S.G. Burston, and A.R. Clarke, (1997). Binding, encapsulation and ejection: substrate dynamics during a chaperonin-assisted folding reaction. *J Mol Biol* **266,** 656-64.
58. Rye, H.S., et al., (1997). Distinct actions of cis and trans ATP within the double ring of the chaperonin GroEL. *Nature* **388,** 792-8.
59. Todd, M.J., G.H. Lorimer, and D. Thirumalai, (1996). Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. *Proc Natl Acad Sci U S A* **93,** 4030-5.
60. Chan, H.S. and K.A. Dill, (1996). A simple model of chaperonin-mediated protein folding. *Proteins* **24,** 345-51.
61. Fink, A.L., (1998). Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Fold Des* **3,** R9-23.
62. Chiti, F., et al., (1999). Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc Natl Acad Sci U S A* **96,** 3590-4.
63. Fandrich, M., M.A. Fletcher, and C.M. Dobson, (2001). Amyloid fibrils from muscle myoglobin. *Nature* **410,** 165-6.
64. Guijarro, J.I., et al., (1998). Amyloid fibril formation by an SH3 domain. *Proc Natl Acad Sci U S A* **95,** 4224-8.
65. Bucciantini, M., et al., (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416,** 507-11.
66. Buard, J. and G. Vergnaud, (1994). Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). *EMBO J.* **13,** 3203-10.
67. van Belkum, A., et al., (1998). Short-sequence DNA repeats in prokaryotic genomes. *Microbiol Mol Biol Rev*. **62,** 275-93.
68. Mitas, M., (1997). Trinucleotide repeats associated with human disease, in Nucleic Acids Res. **25,** 2245-54.
69. Marcotte, E.M., et al., (1999). A census of protein repeats. *J Mol Biol*. **293,** 151-60.

70. Letunic, I., et al., (2002). Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* **30,** 242-244.
71. Groves, M.R. and D. Barford, (1999). Topological characteristics of helical repeat proteins. *Current Opinion in Structural Biology* **9,** 383-389.
72. Main, E.R., S.E. Jackson, and L. Regan, (2003). The folding and design of repeat proteins: reaching a consensus. *Curr Opin Struct Biol*. **13,** 482-9.
73. Andrade, M.A., C. Perez-Iratxeta, and C.P. Ponting, (2001). Protein repeats: structures, functions, and evolution. *J Struct Biol* **134,** 117-31.
74. Peifer, M., S. Berg, and A.B. Reynolds, (1994). A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell* **76,** 789-91.
75. Riggleman, B., E. Wieschaus, and P. Schedl, (1989). Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene. *Genes Dev* **3,** 96-113.
76. Kobe, B., (1999). Autoinhibition by an internal nuclear localization signal revealed by the crystal structure of mammalian importin alpha. *Nat Struct Biol* **6,** 388-97.
77. Andrade, M.A. and P. Bork, (1995). HEAT repeats in the Huntington's disease protein. *Nat Genet* **11,** 115-6.
78. Neuwald, A.F. and T. Hirano, (2000). HEAT repeats associated with condensins, cohesins, and other complexes involved in chromosome-related functions. *Genome Res* **10,** 1445-52.
79. Chook, Y.M. and G. Blobel, (1999). Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. *Nature* **399,** 230-7.
80. Hirano, T., et al., (1990). Snap helix with knob and hole: essential repeats in S. pombe nuclear protein nuc2+. *Cell* **60,** 319-28.
81. Sikorski, R.S., et al., (1990). A repeating amino acid motif in CDC23 defines a family of proteins and a new relationship among genes required for mitosis and RNA synthesis. *Cell* **60,** 307-17.
82. Ponting, C.P., et al., (1999). Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J Mol Biol* **289,** 729-45.
83. Das, A.K., P.W. Cohen, and D. Barford, (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J*. **17,** 1192-9.
84. Main, E.R., et al., (2003). Design of stable alpha-helical arrays from an idealized TPR motif. *Structure (Camb)* **11,** 497-508.
85. Tsukahara, F., et al., (1998). Molecular characterization of the mouse mtprd gene, a homologue of human TPRD: unique gene expression suggesting its critical role in the pathophysiology of Down syndrome. *J Biochem (Tokyo)* **123,** 1055-63.
86. Sohocki, M.M., et al., (2000). Mutations in a new photoreceptor-pineal gene on 17p cause Leber congenital amaurosis. *Nat Genet* **24,** 79-83.
87. Neer, E.J., et al., (1994). The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371,** 297-300.
88. Russell, R.B., P.D. Sasieni, and M.J. Sternberg, (1998). Supersites within superfolds. Binding site similarity in the absence of homology. *J Mol Biol* **282,** 903-18.
89. Madrona, A.Y. and D.K. Wilson, (2004). The structure of Ski8p, a protein regulating mRNA degradation: Implications for WD protein structure. *Protein Sci* **13,** 1557-65.
90. Murzin, A.G., A.M. Lesk, and C. Chothia, (1992). beta-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 beta and 1 alpha and fibroblast growth factors. *J Mol Biol* **223,** 531-43.
91. Takahashi, N., Y. Takahashi, and F.W. Putnam, (1985). Periodicity of leucine and tandem repetition of a 24-amino acid segment in the primary structure of leucine-rich alpha 2-glycoprotein of human serum. *Proc Natl Acad Sci U S A* **82,** 1906-10.
92. Kobe, B. and J. Deisenhofer, (1993). Crystal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature* **366,** 751-6.

93. Tong, Z.B., L.M. Nelson, and J. Dean, (2000). Mater encodes a maternal protein in mice with a leucine-rich repeat domain homologous to porcine ribonuclease inhibitor. *Mamm Genome* **11,** 281-7.
94. Mutai, H., et al., (2000). PAL31, a novel nuclear protein, expressed in the developing brain. *Biochem Biophys Res Commun* **274,** 427-33.
95. Bilder, D. and N. Perrimon, (2000). Localization of apical epithelial determinants by the basolateral PDZ protein Scribble. *Nature* **403,** 676-80.
96. Linhoff, M.W., et al., (2001). Two distinct domains within CIITA mediate self-association: involvement of the GTP-binding and leucine-rich repeat domains. *Mol Cell Biol* **21,** 3001-11.
97. Inohara, N., et al., (1999). Nod1, an Apaf-1-like activator of caspase-9 and nuclear factor-kappaB. *J Biol Chem* **274,** 14560-7.
98. Lux, S.E., K.M. John, and V. Bennett, (1990). Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature* **344,** 36-42.
99. Bork, P., (1993). Hundreds of Ankyrin-Like Repeats in Functionally Diverse Proteins - Mobile Modules That Cross Phyla Horizontally. *Proteins-Structure Function and Genetics* **17,** 363-374.
100. Sedgwick, S.G. and S.J. Smerdon, (1999). The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends in Biochemical Sciences* **24,** 311-316.
101. Michel, F., et al., (2001). Crystal structure of the ankyrin repeat domain of Bcl-3: a unique member of the IkappaB protein family. *Embo J* **20,** 6180-90.
102. Diederich, R.J., et al., (1994). Cytosolic interaction between deltex and Notch ankyrin repeats implicates deltex in the Notch signaling pathway. *Development*. **120,** 473-81.
103. Sherr, C.J. and J.M. Roberts, (1999). CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev* **13,** 1501-12.
104. Ruas, M. and G. Peters, (1998). The p16INK4a/CDKN2A tumor suppressor and its relatives. *Biochim Biophys Acta*. **1378,** F115-77.
105. Penkett, C.J., et al., (2000). Identification of residues involved in the interaction of Staphylococcus aureus fibronectin-binding protein with the (4)F1(5)F1 module pair of human fibronectin using heteronuclear NMR spectroscopy. *Biochemistry* **39,** 2887-93.
106. van Raaij, M.J., et al., (1999). A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401,** 935-8.
107. Binz, H.K., et al., (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J Mol Biol* **332,** 489-503.
108. Luthy, L., M.G. Grutter, and P.R. Mittl, (2002). The crystal structure of Helicobacter pylori cysteine-rich protein B reveals a novel fold for a penicillin-binding protein. *J Biol Chem* **277,** 10187-93.
109. Phelan, P., et al., (2002). Salt bridges destabilize a leucine zipper designed for maximized ion pairing between helices.*Biochemistry* **41,** 2998-3008.

2 Folding of an Ankyrin Repeat Protein

2.1 Introduction

Proteins are often classified on the basis of the conserved sequence motifs that they contain, although the function and structure of the motif might be uncertain. In 1987, Breeden and Nasmyth reported a ~33 residue repeating motif in the sequence of two yeast cell-cycle regulators, Swi6p and Cdc10p, and in the Notch (*Drosophila melanogaster)* and LIN-12 (*Caenorhabditis elegans)* developmental regulators [1]. The discovery of 24 copies of this sequence in the cytoskeletal protein ankyrin led to the naming of this motif as the ankyrin repeat [2]. Ankyrin repeat (AR) proteins carry a wide variety of biological activities and are present in all three super kingdoms including bacteria, archaea, and eukarya, as well as in a number of viral genomes. AR proteins are localized in the nucleus (e.g., IκBα), cytoplasm (e.g., ankyrin), anchored to the membrane (e.g., notch), and secreted into the extracellular space (e.g., black widow spider toxin), which indicates that these proteins can adapt to many different environments. Their biological importance is highlighted by the fact that there are more than 2000 known AR proteins with over 14,000 repeats [3]. Some AR proteins consist solely of ankyrin repeats; others are multi-domain molecules, in which ARs are combined with unrelated structural modules. So far the main known function of AR proteins is to mediate protein-protein interactions. The corresponding complexes are essential for many cellular processes such as cell differentiation, endocytosis, transcription regulation, and cell cycle control.

A number of ankyrin repeat proteins have been linked to human diseases. For example, the mammalian cell cycle inhibitor p16 is encoded by the *INK4a* gene, which is a frequent site of mutations that are strongly associated with cancer [4]. A mutation in the ankyrin repeats of human Notch3 is associated with a hereditary adult-onset condition causing stroke and dementia [5].

2.1.1 Structural organization of ankyrin repeats (AR)

Almost ten years after the discovery of the AR motif, the first three-dimensional structure of an ankyrin repeat molecule, 53BP2 bound to the p53 cell-cycle tumour suppressor was determined [6]. This and subsequent structures have shown that AR proteins are composed of stacked ARs,

consisting typically of 33 amino acid residues, each forming two antiparallel helices with a β-hairpin/loop region projecting outward from the helices at a 90° angle. Because the set of helices that face away from the β-hairpin/loop region are slightly longer, the stacked repeats form a concave L-shaped structure (Fig. 2.1). The inter-repeat interface consists of hydrophobic interactions stabilizing the interface between the helices of consecutive repeats in addition to a hydrogen bonding network connecting the β-hairpin/loop region. Usually, four to six repeats assemble into domains [7], but the crystal structure of ankyrin R consisting of 12 ARs in a single domain [8] and the presence of 29 consecutive repeats in a single protein [9] indicate that there is virtually no limit to the number of repeats that can fold in one AR domain.

Figure 2.1 The structure of the ankyrin domain of the hepatocellular carcinoma-associated protein gankyrin (PDB code: 1UOH). The ribbon representation shows five internal ankyrin repeats along with flanking N- (blue) and C-termini (red) repeats and antiparallel β-hairpins [10].

To date, the structures of 13 naturally occurring and three designed ankyrin repeat proteins have been solved. These structures closely resemble one another despite their different cellular functions, supporting the role of the ankyrin repeat as a versatile scaffold for protein–protein interactions. Indeed, it appears that the ankyrin repeat motif is defined by its fold rather than by its function, since no specific sequence or structure motif that is universally recognized by AR has been identified so far.

Consensus design of ankyrin repeats (AR) – The stacked L-shaped structure of ankyrin repeats has evolved as a scaffold for various protein-protein interactions. The motif may serve as a template or building block for protein engineering and design studies. Due to the large number of currently available sequences, it is an especially attractive model for consensus-based protein design.

Recent efforts at designing ankyrin repeat proteins using sequence-based strategies have been successful [11, 12]. Mosavi et al. performed statistical analyses on ~4400 ankyrin repeat sequences deposited in the PFAM database to create a consensus sequence, or idealized ankyrin repeat [11]. Their analyses consisted of classifying the conservation-level of each position based on the amino acid occurrence at that position, the amino acid distribution over the entire sequence, and the secondary structure elements corresponding to each position. The ensuing sequence was used to construct proteins consisting of one, two, three, or four identical repeats (Fig. 2.2).

Figure 2.2 The structure of a consensus designed ankyrin repeat protein containing four identical sequence repeats by Mosavi et al. [11]. Individual repeats are colored differently from N- to C-terminus (PDB code: 1N0R).

Binz et al. used a different strategy. To create a consensus ankyrin sequence, they started with the statistical analysis of 229 repeats from the SMART database and included information from 10 existing high-resolution structures of ankyrin repeat proteins [12]. They designed a consensus AR consisting of fixed framework residues and randomized interacting residues. N- and C- terminal "capping" repeats derived from the structure of GABP-β were incorporated into the design. The role of the capping repeats is to shield the hydrophobic core of the first and last internal AR from the solvent. To form a potential interaction surface, positions 2, 3, 5, 13, 14, and 33 (numbering for first repeat) in the shorter helix and β-hairpin region were randomized to contain any amino acid except glycine, proline, or cysteine. Varying numbers of internal repeats were inserted between the N- and C-terminal caps. As a result, designed AR protein libraries of distinct repeat numbers were obtained. The nomenclature is N2C, N3C, N4C, etc., where N and C denote the caps and digits denote the number of consensus repeats. The X-ray crystal structure of the N3C library member E3_5 (a five-repeat library member) is shown in Fig. 2.3. The structure is well-packed adopting a regular ankyrin repeat fold [13]. Using this consensus design, Binz et al. have

generated combinatorial libraries of AR proteins of varying repeat numbers with diversified binding surfaces. Successful selection of library members that bind with high affinity to *Escherichia coli* maltose binding protein (MBP) and to two eukaryotic mitogen-activated protein kinases (MAPKs) was achieved recently [14].

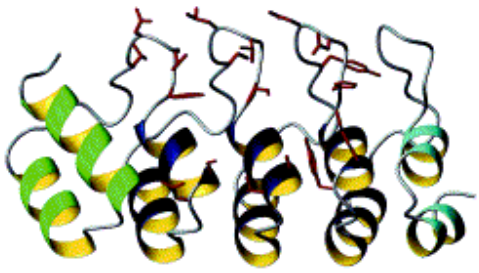


Figure 2.3 Crystal structure of E3_5 (PDB entry: 1MJ0), a designed ankyrin repeat protein consisting of five repeats (N3C). Residues represented as red sticks are randomized positions that build a large potential interaction surface. The N and C-terminal capping repeats are in green and the internal repeat modules are in blue [13].

2.1.2 Folding studies of natural ankyrin repeats

Until recently folding studies have focused mainly on globular proteins. Ankyrin repeats represent a class of proteins with an elongated structure and little or no long-range contacts. From a folding perspective, the modular nature of ankyrin repeat domains presents a number of interesting questions. Unlike the packing of globular protein domains, the linear packing of the repeat modules in AR proteins implies that local, regularly repeating packing interactions are very important and may dominate the thermodynamic stability and the folding mechanism. One is tempted to assume that AR proteins fold in a modular, multistate reaction controlled by short-range interactions. Interestingly, however, folding studies on a number of naturally occurring ankyrin repeat proteins show that the ankyrin repeat domain follows an apparent two-state folding pathway. The ankyrin domain of *Drosophila* Notch (a signaling protein with 7 AR) and two variants with one or two C-terminal repeats deleted fold in a two-state manner [15]. The ankyrin domain of Notch has an unfolding free energy of ΔG_{H2O} = 8.03 kcal mole^{-1}. The human AR proteins p16 (four AR) and p19 (five AR), both members of the INK4 tumor suppressor family,

show three refolding phases at low urea concentrations and one unfolding phase [16, 17]. Some truncation studies of AR proteins suggest that four to six repeats are necessary to form a stably folded structure, and that a single repeat is not capable of folding [18]. On the other hand, truncation studies of the AR protein p16 demonstrate that the C-terminal two repeats form an autonomously folded unit and suggest that two repeats are the minimum folding unit of the ankyrin repeat [19].

2.1.3 Stability of designed ankyrin repeats

Thermal melting and equilibrium unfolding studies have been performed on randomly selected members of the designed ankyrin consensus library by Binz et al. [12, 13]. All the selected proteins had well-folded structures and were stable. Equilibrium unfolding studies with GdmCl of designed four, five and six AR proteins showed cooperative and reversible unfolding with ΔG_{H2O} between 9.5 and 21 kcal mol^{-1}. When compared to naturally occurring AR proteins, the thermodynamic stability of the designed repeat proteins is quite high. Thermal melting midpoint temperatures (T_m) are in the range of 66 °C to 85 °C depending on the number of repeats [12]; naturally occurring AR proteins have T_m of 50 °C or less [15, 20]. Obviously the designed AR proteins are tightly folded and very stable.

To further characterize the behavior of the “idealized” consensus AR used in the library of Binz et al., we have investigated the stability and folding of one randomly chosen member of the N1C library consisting of a single consensus AR flanked by terminal capping repeats.

2.2 References

1. Breeden, L. and K. Nasmyth, (1987). Similarity between cell-cycle genes of budding yeast and fission yeast and the Notch gene of Drosophila. *Nature* **329,** 651-4.
2. Lux, S.E., K.M. John, and V. Bennett, (1990). Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature* **344,** 36-42.
3. Letunic, I., et al., (2002). Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res*. **30,** 242-4.
4. Ruas, M. and G. Peters, (1998). The p16INK4a/CDKN2A tumor suppressor and its relatives. *Biochim Biophys Acta*. **1378,** F115-77.
5. Joutel, A., et al., (1996). Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. *Nature* **383,** 707-10.
6. Gorina, S. and N.P. Pavletich, (1996). Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science* **274,** 1001-5.
7. Jacobs, M.D. and S.C. Harrison, (1998). Structure of an IkappaBalpha/NF-kappaB complex. *Cell* **95,** 749-58.
8. Bork, P., (1993). Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins* **17,** 363-74.
9. Michaely, P., et al., (2002). Crystal structure of a 12 ANK repeat stack from human ankyrinR. *EMBO J*. **21,** 6387-96.
10. Walker, R.G., A.T. Willingham, and C.S. Zuker, (2000). A Drosophila mechanosensory transduction channel. *Science* **287,** 2229-34.
11. Mosavi, L.K., D.L. Minor, Jr., and Z.Y. Peng, (2002). Consensus-derived structural determinants of the ankyrin repeat motif. *Proc Natl Acad Sci U S A*. **99,** 16029-34.
12. Binz, H.K., et al., (2003). Designing repeat proteins: well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J Mol Biol*. **332,** 489-503.
13. Kohl, A., et al., (2003). Designed to be stable: crystal structure of a consensus ankyrin repeat protein. *Proc Natl Acad Sci U S A*. **100,** 1700-5.
14. Binz, H.K., et al., (2004). High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat Biotechnol*. **22,** 575-82.
15. Zweifel, M.E. and D. Barrick, (2001). Studies of the ankyrin repeats of the Drosophila melanogaster Notch receptor. 2. Solution stability and cooperativity of unfolding. *Biochemistry* **40,** 14357-67.
16. Tang, K.S., et al., (1999). Stability and folding of the tumour suppressor protein p16. *J Mol Biol*. **285,** 1869-86.
17. Zeeb, M., et al., (2002). Protein folding and stability of human CDK inhibitor p19(INK4d). *J Mol Biol*. **315,** 447-57.
18. Michaely, P. and V. Bennett, (1993). The membrane-binding domain of ankyrin contains four independently folded subdomains, each comprised of six ankyrin repeats. *J Biol Chem*. **268,** 22703-9.
19. Zhang, B. and Z. Peng, (2000). A minimum folding unit in the ankyrin repeat protein p16(INK4). *J Mol Biol*. **299,** 1121-32.
20. Mosavi, L.K., S. Williams, and Z.Y. Peng Zy, (2002). Equilibrium folding and stability of myotrophin: a model ankyrin repeat protein. *J Mol Biol*. **320,** 165-70.

2.3 Folding of a designed simple ankyrin repeat protein

V. Sathya Devi, H. Kaspar Binz, Michael T. Stumpp, Andreas Plückthun, Hans Rudolf Bosshard and Ilian Jelesarov

Protein Science **13**, 2864-2870 (2004)

Folding of a designed simple ankyrin repeat protein

V. SATHYA DEVI, H. KASPAR BINZ, MICHAEL T. STUMPP, ANDREAS PLÜCKTHUN, HANS RUDOLF BOSSHARD, AND ILIAN JELESAROV

Department of Biochemistry, University of Zurich, CH-8057 Zurich, Switzerland

(RECEIVED June 17, 2004; FINAL REVISION August 2, 2004; ACCEPTED August 5, 2004)

Abstract

Ankyrin repeats (AR) are 33-residue motifs containing a β-turn, followed by two α-helices connected by a loop. AR occur in tandem arrangements and stack side-by-side to form elongated domains involved in very different cellular tasks. Recently, consensus libraries of AR repeats were constructed. Protein E1_5 represents a member of the shortest library, and consists of only a single consensus repeat flanked by designed N- and C-terminal capping repeats. Here we present a biophysical characterization of this AR domain. The protein is compactly folded, as judged from the heat capacity of the native state and from the specific unfolding enthalpy and entropy. From spectroscopic data, thermal and urea-induced unfolding can be modeled by a two-state transition. However, scanning calorimetry experiments reveal a deviation from the two-state behavior at elevated temperatures. Folding and unfolding at 5°C both follow monoexponential kinetics with $k_{folding} = 28\ sec^{-1}$ and $k_{unfolding} = 0.9\ sec^{-1}$. Kinetic and equilibrium unfolding parameters at 5°C agree very well. We conclude that E1_5 folds in a simple two-state manner at low temperatures while equilibrium intermediates become populated at higher temperatures. A chevron-plot analysis indicates that the protein traverses a very compact transition state along the folding/unfolding pathway. This work demonstrates that a designed minimal ankyrin repeat protein has the thermodynamic and kinetic properties of a compactly folded protein, and explains the favorable properties of the consensus framework.

Keywords: ankyrin repeat; calorimetry; protein design; protein folding; protein stability

A large number of protein classes are built on the modularity principle from homologous structural blocks (Marcotte et al. 1999; Letunic et al. 2002). Repeat proteins consist of a number of structurally identical motifs usually arranged in tandem, which stack to form elongated or supercoiled domains (Groves and Barford 1999). Repeating modules are typically 20–40 residues long, and contain secondary structure elements folding in a variety of topologies (Main et al. 2003). The linear assemblage of complementary repeats results in a relatively simple and robust scaffold, which is maintained by the regular repetition of hydrophobic contacts and hydrogen bonds. The architecture of stacked elements implies that stabilizing contacts are either within one repeat or between directly adjacent repeats, but there are no contacts between residues distant in sequence. Repeat proteins have adapted to different environments and promote ligand recognition through an array of variable functionalities in the side chains and sometimes in the main chain.

Ankyrin repeat proteins occur in virtually all species, even though the majority is found in eukaryots, and are involved in a wide range of cellular tasks, ranging from transcriptional regulation to cytoskeleton organization (Bork 1993). The ankyrin repeat (AR) is a 33-residue L-shaped motif, which contains two antiparallel α-helices connected by a short loop (Sedgwick and Smerdon 1999). The consecutive repeats stack in parallel and are joined by β-hairpins forming the base of the L. Four to six repeats are typical, but as many as 12 AR (Michaely et al. 2002) and 29 AR (Walker et al. 2000) in a single domain have been reported.

Due to their modular structure, repeat proteins in general, and AR in particular, are very attractive experimental ob-

Reprint requests to: Ilian Jelesarov, Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland; e-mail: iljel@bioc.unizh.ch; fax: ++41-1-635-6805.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.04935704.

jects both for testing our understanding about sequence–structure–stability–function relationships in proteins, and for developing molecular tools for biotechnological applications like, for example, specific molecular recognition (Binz et al. 2004; Forrer et al. 2004). Unlike the packing of globular protein domains, the linear packing of the repeat modules in AR proteins implies that local, regularly repeating packing interaction patterns are very important or even dominating the thermodynamic stability and folding mechanism (McDonald and Peters 1998). Indeed, the analysis of crystal structures has demonstrated that hydrophobic interactions between the helices within a single AR are not well optimized, while hydrophobic packing is tighter at interrepeat interfaces (Kohl et al. 2003). This peculiarity has prompted studies aimed at the elucidation of the thermodynamic stability as a function of the repeat number within a single AR domain consisting of several repeats (Zweifel and Barrick 2001b; Binz et al. 2003), the cooperative behavior and its limits (Bradley and Barrick 2002), the identification of minimal folding units (Zhang and Peng 2000; Mosavi et al. 2002), the thermodynamic consequences of mutations (Mosavi and Peng 2003; Zweifel et al. 2003), and the folding mechanism (Tang et al. 1999, 2003; Zeeb et al. 2002).

More recently, the design of novel AR has been reported (Mosavi et al. 2002; Binz et al. 2003, 2004). The two successful design strategies are both based on sequence data base analysis and identification of residues maintaining the AR fold. A multiple sequence alignment and statistical analysis was used to calculate the probability of amino acid usage at each position of AR (Mosavi et al. 2002). The successful application of a novel design strategy to construct combinatorial AR protein libraries to select specific binders was reported (Binz et al. 2003, 2004; Kohl et al. 2003). Sequence consensus analysis refined by structural considerations has led to the design of a 33 amino acid AR module in which seven positions are randomized to obtain AR libraries (Binz et al. 2003; Forrer et al. 2003; Kohl et al. 2003). To render repeat proteins soluble and monomeric the exposed hydrophobic faces of the terminal repeats are shielded by capping repeats. Library members containing two to four internal repeats flanked by N- and C-terminal capping motifs are soluble, do not oligomerize and display high thermodynamic stability (Binz et al. 2003). To characterize the behavior of the "idealized" consensus AR that are used in the library, here we investigate the stability and folding of one randomly chosen member of the smallest library consisting of a single consensus AR flanked by terminal capping repeats.

Results

The simplest AR library proteins consist of a single central consensus repeat flanked by N-terminal and C-terminal capping repeats (Fig. 1). Protein E1_5 was expressed in soluble

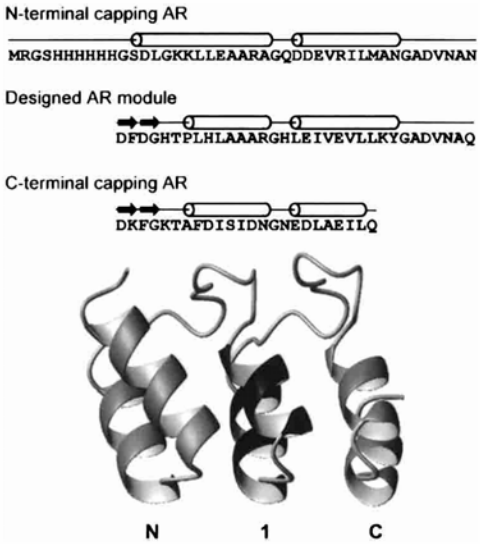


Figure 1. The E_5 protein. After the N-terminal His tag, the sequence of the three ankyrin repeats are shown on separate lines. For orientation, the secondary structure elements are indicated above the sequence. Cylinders, α-helices; arrows, β-turns. In the structural model, the flanking capping repeats are shown in light gray, while the central consensus repeat is in dark gray. The N1C AR protein model was generated using homology modeling with Insight II (Accelrys) and the crystal structures of GABPb1 (PDB ID: 1AWC; Batchelor et al. 1998), E3_5 (1MJ0; Kohl et al. 2003), 3ANK (1N0Q; Mosavi et al. 2002), and 4ANK (1N0R; Mosavi et al. 2002) as templates. The picture was created using MOLMOL (Koradi et al. 1996).

form in the cytoplasm of *Escherichia coli* at high yield (80 mg purified protein per one liter of culture). The purified protein is monomeric at the concentrations and under the experimental conditions used in this study, as evidenced by gel filtration and multiangle light scattering (not shown). The far-UV CD spectrum has the typical spectral signature of naturally occurring and designed AR (not shown).

Thermal unfolding

Unfolding of E1_5 at pH 7.0 was monitored by following the temperature-induced changes in ellipticity at 222 nm. The melting traces were superimposable with protein concentrations between 10 and 150 μM, thus demonstrating that no aggregation takes place in the temperature range studied. Thermal unfolding was reversible to >95% when the protein was heated to 65°C. Heating above that temperature reduced the reversibility due to a time-dependent process, possibly asparagine deamidation occurring at elevated temperatures in the deamidation-prone Asn-Ala and Asn-Gly sequences (Robinson and Robinson 2001). Therefore, CD data were analyzed only between 3°C and 65°C. The unfolding transition is sigmoidal, without indication for consecutive unfolding steps, yet it is quite broad (Fig. 2A).

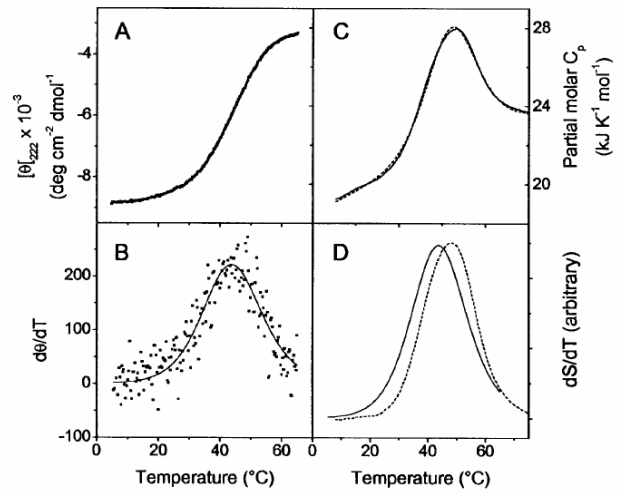


Figure 2. Thermal melting of E1_5 at pH 7.0, followed by CD spectroscopy and DSC. Temperature-induced changes in MRE_{222} are shown in *A*. Protein concentration was 30 μM. (*B*) The same data in the derivative mode (symbols). The best fit according to equations 1–3 with T_m = 44.2°C and $\Delta H_{vH,CD}$ = 130 ± 10 kJ mole^{-1} is visualized by the continuous line. (*C*) Partial molar heat capacity changes upon heating at 1° min^{-1}. Protein concentration was 150 μM. The experimental data are shown as a continuous line. The broken line is the best fit for a two-state unfolding model determined with T_m = 48°C, ΔH_{fit} = 136 ± 5 kJ mole^{-1} and ΔC_p = 2.0 ± 0.2 kJ K^{-1} mole^{-1}. (*D*) Superposition of the excess heat capacity obtained by calorimetry (dashed line) and the temperature derivative of MRE_{222} (continuous line) is shown. To facilitate the comparison, both functions were arbitrarily scaled.

Because the posttransitional portion of the signal change is not well defined if heating proceeds to 65°C only, the data were analyzed in the differential mode by the combined equations 1–3 (Fig. 2B; John and Weeks 2000). The melting curves are compatible with a two-state unfolding transition with midpoint T_m = 44.2°C and van't Hoff enthalpy $\Delta H_{vH,CD}$ = 130 ± 10 kJ mole^{-1}.

The thermal stability and unfolding of E1_5 were further characterized by differential scanning calorimetry. In the native region between 5° and 25°C, the partial specific heat capacity is a linear function of temperature and increases with a $\partial C_p/\partial T = (6.8 \pm 1.0) \times 10^{-3}$ J K^{-2} g^{-1}. Overall, the temperature dependence of the partial molar heat capacity can be reasonably modeled by a two-state conformational transition (Fig. 2C). However, closer analysis reveals a small but significant deviation from the two-state model. Notably, the maximum of the heat capacity function appears at 48°C, higher than the T_m detected by CD spectroscopy (44°C; Fig. 2D), although the best fit of the two-state unfolding model to the data was obtained with $\Delta H_{fit,DSC}$ = 136 ± 7 kJ mole^{-1}, which is identical within error with $\Delta H_{vH,CD}$. The calorimetric, model-independent enthalpy, ΔH_{cal}, amounts to 145 ± 5 kJ mole^{-1}, while the effective van't Hoff enthalpy, $\Delta H_{vH,DSC}$, obtained by analysis of the shape of the heat absorption peak, is 129 ± 10 kJ mole^{-1}. Thus, the ratio $\Delta H_{vH,DSC}/\Delta H_{cal}$ is 0.89, indicative of a population of intermediate states in the transition zone. The data are best described by an unfolding heat capacity increment ΔC_p = 2.0 ± 0.2 kJ K^{-1}mole^{-1}. From thermal melting data, according to the Gibbs-Helmholtz equation, the free energy of unfolding of E1_5 at 5°C is ΔG_{th} = 11 ± 1 kJ mole^{-1}, higher than ΔG from urea induced unfolding (see below).

Urea-induced unfolding

The stability of E1_5 at 5°C was assessed from isothermal urea-induced unfolding experiments by following the change in ellipticity at 222 nm. As in thermal unfolding, the unfolding curve is rather broad, but the data can be modeled with a two-state transition between native and unfolded protein with a midpoint at 1.8 M urea (Fig. 3). The linear extrapolation procedure yields $\Delta G_{ur,H_2O}$ = 7.7 ± 0.8 kJ mole^{-1} at 5°C and 0 urea. The urea dependence of ΔG, $m_{eq} = -\partial\Delta G_{urea}/\partial$[urea], is 5.7 ± 0.5 kJ mole^{-1} M^{-1}. Experiments were performed also at higher temperatures. The midpoint of unfolding shifts to slightly lower denaturant concentrations. Unfortunately, the pretransition portion of the unfolding curve is not well defined above 5°C, which precluded accurate determination of ΔG_{unf} at higher temperatures. However, according to our semiquantitative estimates, the combination of $[urea]_{1/2}$ and m_{eq} is such that the decrease in stability is less than 1 kJ mole^{-1} between 5° and 20°C.

Folding/unfolding kinetics

The rate of folding and unfolding of E1_5 was studied by following the time course of secondary structure formation/disruption after rapid dilution from or into urea at 5°C. Representative kinetic traces are shown in Figure 4. Folding and unfolding are both described precisely by single exponential phases at all urea concentrations tested. Throughout

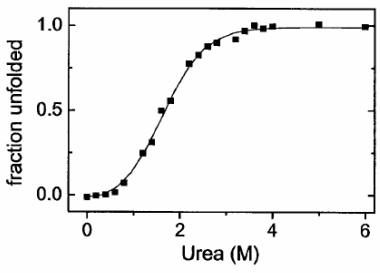


Figure 3. Equilibrium urea-induced unfolding of E1_5 at 5°C followed by CD spectroscopy. The symbols are the experimental data presented as fraction unfolded protein. The continuous line is calculated by the linear extrapolation model for a two-state transition. Protein concentration was 25 μM.

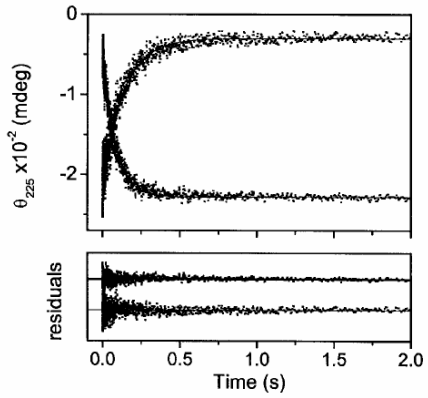


Figure 4. Representative kinetic traces describing refolding and unfolding of E1_5 at 5°C followed by CD stopped-flow. The lines are best fits according to a single exponential function. The residuals of the fits are shown in the *lower* panels. Final protein concentration was 25 μM. Final urea concentration was 0.45 M for refolding (*downward* trace) and 5.3 M for unfolding (*upward* trace), respectively.

the entire range of urea concentrations the amplitudes of the kinetic traces account for >95% of the spectral differences between the folded and unfolded state measured at equilibrium. It follows that there are no kinetic phases observable by CD that are hidden in the dead time of the measurement. No curvature in the refolding and unfolding limbs of the Chevron plot is detected in the range of urea concentrations studied (Fig. 5). According to equation 4, the data set is consistent with $k_{f,H_2O} = 28 \pm 2$ sec^{-1} and $k_{unf,H_2O} = 0.9 \pm 0.1$ sec^{-1}. The vertex of the Chevron plot is at 1.9 M urea. From kinetic data, the free energy of unfolding at 5°C is $\Delta G_{kin,H_2O} = -RT\ln(k_{unf,H_2O}/k_{f,H_2O}) = 7.9 \pm 0.9$ kJ mole^{-1}, $m_f = -1.85 \pm 0.2$ M^{-1}, $m_{unf} = 0.43 \pm 0.04$ M^{-1} and $m_{kin} = RT(|m_f| + |m_{unf}|) = 5.3 \pm 0.5$ kJ mole^{-1} M^{-1}.

Discussion

Peptides corresponding to the length of a single AR and being compactly packed and stable have not been reported so far. The reason for the intrinsic instabiliy of an isolated AR is that it is too small to form a well-developed hydrophobic core and cooperative accumulation of interrepeat contacts provides substantial stabilization in a folded domain of several AR. Two repeats can be sufficient to form a stable folding unit. It was demonstrated that the third and fourth AR of the tumor supressor p16^{INK4} fold autonomously and cooperatively (Zhang and Peng 2000). Nevertheless, naturally occurring proteins very rarely contain less than three AR. Cooperative folding into parallel stacks of only a few repeats faces the problem of how to shield the hydrophobic core from the solvent, because the complementary surface of the tandem repeats is largely hydrophobic. In nature, specialized terminal repeats exposing one predominantly hydrophilic face to the solvent are often found to terminate the internal hydrophobic stack of AR domains. In the present design of consensus AR libraries (Binz et al. 2003; Forrer et al. 2003; Kohl et al. 2003) special capping repeats are used facilitating the correct folding of library members with a varying number of internal repeats. The success of this strategy has been demonstrated by proteins containing two to four AR between the caps (Binz et al. 2003; Forrer et al. 2003; Kohl et al. 2003). In contrast, the two-repeat consensus construct originally designed by Mosavi et al. (2ANK) is partially folded under some conditions but is not monomeric (Mosavi et al. 2002). Leucine to arginine substitutions on the surface of 2ANK allowed the partially folded protein to assume a fully folded conformation (Mosavi and Peng 2003), yet at the expense of thermodynamic stability, possibly due to local repulsions between like charges.

In the present study we show that the simplest possible library member containing a single internal repeat and two capping repeats is stably folded. The E1_5 protein is soluble and monomeric. MRE_{222} (mean residue ellipticity at 222 nm) is −8700° cm^{-2} dmole^{-1}. This helical content is slightly lower than what has been observed for library members containing two to four internal repeats, but it is comparable with the very stable four-repeat domain designed by Mosavi et al. (2000) and the *Drosophila* Notch protein containing five, six, or seven repeats (Zweifel and Barrick 2001a). Calorimetric data provide strong support that the protein is compactly folded. The specific unfolding enthalpy, 1.6 kJ (mole res)$^{-1}$, the specific unfolding entropy, 5.1 J K^{-1} (mole res)$^{-1}$ at 48°C, as well as the temperature slope of the heat capacity of the folded protein are typical for globular proteins domains (Gomez et al. 1995; Makhatadze and Privalov 1995). (In the calculation, the N-terminal his-tag 12 residues [Fig. 1] are not taken into consideration, because they are unfolded and influence only negligibly the measured thermodynamic parameters.) These observations indicate that the enthalpy and entropy factors determining E1_5 stability are balanced similarly to globular proteins.

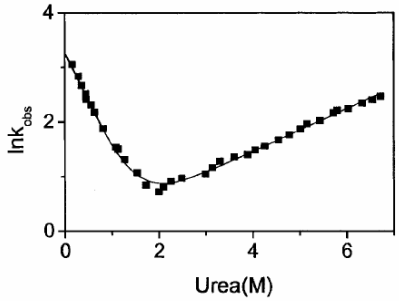


Figure 5. Urea dependence of the rate constants for refolding (*left* limb) and unfolding (*right* limb). The observed relaxation constants are presented as symbols. The continuous line was calculated according to equation 4.

The unfolding transitions induced by heat and urea are relatively broad, implying that the cooperativity of tertiary structure consolidation/disruption is low, as is usual for proteins of small size (Fig. 2). Because the effective unfolding enthalpies measured by CD spectroscopy and calorimetry are lower than the model-independent calorimetric estimate, intermediate(s) become(s) populated at higher temperatures. The midpoint of thermal unfolding is shifted to higher temperatures when the transition is monitored by the changes in partial molar heat capacity (Fig. 2D). Therefore, it is likely that melting of the α-helices precedes disruption of gross packing interactions upon temperature increase, and the intermediate(s) represent(s) a relatively compact species without pronounced helical content. A folding intermediate with the same overall structural features has been detected in folding of the tumor supressor p16 at 25°C (Tang et al. 1999). Statistical thermodynamic modeling of the heat capacity function to obtain the thermodynamic characteristics of the intermediate state(s) was not possible because the population of that state(s) is relatively low. Interestingly, the combined data collected at 5°C are consistent with a two-state behavior. Refolding and unfolding are both mono-exponential (Fig. 4). There is no evidence for a "roll-over" at strongly native conditions (Fig. 5). The free energy of unfolding measured at equilibrium using the two-state approximation is identical within error with the free energy calculated from kinetic data. Kinetic and equilibrium *m*-values also agree well. Possibly, there is a subtle, temperature-dependent change in the folding mechanism from simple two-state at low temperatures to a more complicated mechanism involving intermediate(s) at higher temperatures. Alternatively, the intermediary state(s) escape(s) detection at low temperatures.

In view of the slight differences in the unfolding at low and high temperatures, the different estimates of ΔG_{unf} of E1_5 obtained from thermal and isothermal unfolding data come as no surprise. Possibly, the stability measured directly at 5°C, $\Delta G_{ur,H_2O}$, is more reliable than ΔG_{th} extrapolated according to the Gibbs-Helmholtz equation using thermal melting data, because extrapolation is over a large temperature range and neglects the presence of intermediates. On the other hand, if significant tertiary contacts are retained in the absence of α-helical structure, stability might have been underestimated from both equilibrium and kinetic CD data collected at 5°C.

The protein is marginally stable, 8–11 kJ mole^{-1}. However, construct E2_5 having two internal repeats of the same sequence as E1_5 is dramatically more stable, displaying an increase in T_m of ~30°C, and a sixfold increase in stability compared to E1_5 (Binz et al. 2003). In comparison, the autonomously folded two-repeat fragment of p16^{INK4} has a $\Delta G = 7.1 \pm 1.7$ kJ mole^{-1} (Zhang and Peng 2000), while the full-length four repeat long p16^{INK4} has a ΔG ~13 kJ mole^{-1} (Tang et al. 1999) and the *Drosophila* Notch protein with 6 repeats has $\Delta G = 10 \pm 1.5$ kJ mole^{-1} (Zweifel and Barrick 2001b). Taken together, these observations strengthen the contention that engineered consensus AR proteins are more stable than AR proteins occurring in nature, as recently discussed (Main et al. 2003; Forrer et al. 2004).

E1_5 unfolds and refolds following a single exponential phase when observing the changes in ellipticity at 225 nm. Within the precision of the measurements, both $\ln k_f$ and $\ln k_{unf}$ are linear functions of the urea concentration. Down to 0.15 M urea we do not observe "roll-over" in the refolding limb of the Chevron plot, which would be indicative of a urea-sensitive folding intermediate. This is consistent with an apparent reversible two-state folding of the "idealized" small AR protein. In contrast, the human CDK inhibitor p19INK4d and the tumor supressor p16 exhibit multiphasic folding kinetics in CD and fluorescence stopped-flow experiments (Tang et al. 1999; Zeeb et al. 2002). An intermediate has been postulated for the tumor supressor p16 (Tang et al. 1999). For E1_5, the existence of a folding intermediate accumulating at urea concentrations below 0.15 M, or appearing even at higher urea concentrations, yet not detectable by CD, cannot be completely ruled out. However, the good correspondence between equilibrium and kinetic ΔG and *m*-values argues against this possibility.

Despite its small size, some kinetic properties of E1_5 are very similar to those of the tumor supressor protein p16 containing four AR (Tang et al. 1999). The refolding rate of E1_5 extrapolated to zero urea is 28 sec^{-1} at 5°C. Refolding of p16 proceeds through an intermediate which accumulates very rapidly and interconverts to the native state in the rate-limiting step with a rate constant of 33 sec^{-1} at 25°C. These rates are relatively slow. Also, the unfolding rates are very similar: 0.9 sec^{-1} at 5°C for E1_5 and 0.8 sec^{-1} at 25°C for p16, indicating that the proteins unfold by crossing a similar energy barrier. Furthermore, both proteins exhibit a shallow unfolding limb in their Chevron plots. From the Tanford's ratio, $\beta_T = m_{unf}/(|m_{ref}| + |m_{unf}|)$, it can be concluded that E1_5 and p16 traverse a very compact rate-limiting high-energy state along the refolding/unfolding pathway. This transition state is 89% and 84% native-like with respect to its overall surface exposure for p16 and E1_5, respectively. At least some of these common biophysical features may be the consequence of the AR architecture and the similar thermodynamic stabilities of the two proteins.

In conclusion, the presented data demonstrate that E1_5, containing a central consensus AR flanked by capping repeats is an autonomously folding domain. Folding appears to be a simple two-state process at low temperatures, while equilibrium intermediates become populated at higher temperatures. The protein folds via a very compact transition state. The kinetic stability is low but equilibrium experiments predict a marked increase with the increasing number

of repeats (Binz et al. 2003). With the characterization of the smallest library member the stage is now set for systematic studies of the biophysical properties of AR domains consisting of identical internal repeats as a function of the repeat number.

Materials and methods

Protein expression and purification

The construction and cloning of designed AR protein libraries is described in detail elsewhere (Binz et al. 2003). The sequence of the library member E1_5 is shown in Figure 1. The protein was expressed in the soluble form in XL1-Blue *E. coli* at 37°C after induction with 1 mM IPTG. After 4 h, the cells were harvested by centrifugation, resuspended in 50 mM Tris-HCl, 500 mM NaCl (pH 8.0) and sonicated. The lysate was centrifuged and glycerol (10% final concentration) and imidazole (20 mM final concentration) were added to the resulting supernatant. The protein was purified over a Ni-nitrilotriacetic acid column (2.5 mL column volume) according to the manufacturer's instructions (Qiagen). The purity was checked by SDS-PAGE, and the identity of the protein was verified by mass spectroscopy. The protein was extensively dialyzed against the working buffer (see below) and protein concentrations were determined by UV spectroscopy using ε_{280} = 1280 cm^{-1} M^{-1} (Edelhoch 1967).

Buffers and chemicals

All chemicals were of the highest grade available, and were used without further purification. All experiments were performed in a buffer cocktail containing 7.5 mM each of boric acid, citric acid, and phosphoric acid, 100 mM KCl (pH 7.0). For experiments in urea, the denaturant was added before the pH adjustment. Urea concentrations were determined by measuring the refractive index.

Circular dichroism (CD) spectroscopy

Experiments were performed on a J-715 instrument (Jasco Ltd.) equipped with a computer-controlled water bath, using cylindrical jacketed cuvettes of 1 mm optical path length. Spectra were recorded three times between 200 and 250 nm at scanning rate of 5 nm min^{-1}. Thermal melting curves were recorded by continuous heating at 1° min^{-1}. Data points (ellipticity at 222 nm) were collected every 10 sec. Reversibility was determined from the recovery of the mean residue ellipticity (MRE_{222}) after cooling. Thermal melting curves were analyzed according to (John and Weeks 2000):

$$\frac{\partial\theta_{222}}{\partial T} = A \cdot \frac{\Delta H_m}{RT^2} \cdot f_u \cdot (1 - f_u) \qquad (1)$$

where A is a scaling constant, R is the gas constant, and ΔH_m is the van't Hoff enthalpy at T_m. The fraction of unfolded protein, f_u, is given by:

$$f_u = \frac{K_u(T)}{1 + K_u(T)} \qquad (2)$$

and the equilibrium unfolding constant, $K_u(T)$ is calculated with the van't Hoff expression:

$$K_u(T) = exp\left[\frac{\Delta H_m}{R} \cdot \left(\frac{1}{T_m} - \frac{1}{T}\right)\right] \qquad (3)$$

For measuring of urea melting curves, 25 μM protein was incubated overnight at the corresponding urea concentrations. The signal was averaged over 3 min after thermal equilibration. Urea-induced equilibrium unfolding experiments were analyzed by nonlinear least-squares regression according to well-established procedures (Milev et al. 2003).

Differential scanning calorimetry (DSC)

The temperature dependence of the heat capacity was determined with the VP-DSC calorimeter (MicroCal LLC) at heating rate of 1° min^{-1}. Details on the performance of the instrument are given elsewhere (Plotnikov et al. 1997). After subtraction of the buffer versus buffer baseline, the data were transformed to partial specific heat capacity using a partial specific volume of 0.715 cm^{-3} g^{-1} calculated from the amino acid sequence (Makhatadze et al. 1997). The data were analyzed by nonlinear least-squares regression using the program CpCalc 2.1 (Applied Thermodynamics) or in-house scripts written for NLREG (Phillip Sherod) utilizing thermodynamic modeling as described previously (Milev et al. 2003).

Stopped-flow kinetics

Kinetic experiments were performed with the π*-180 instrument (Applied Photophysics). The dead time was 1–2 msec, and the optical path length was 10 mm. Refolding was initiated by mixing one volume of buffered protein solution (250 μM) containing 4–5 M urea with 10 or 25 volumes of buffer, or with buffer containing various concentrations of the denaturant. Unfolding rates were measured by 1:10 dilution of the protein into solutions containing final urea concentrations >2.5 M. The detection wavelength was 225 nm and the slits were set to 4 mm. Ten to 15 firings were averaged for each kinetic trace. The data were analyzed with the software provided by the manufacturer. The Chevron plot was analyzed by the following equation (Fersht 1999):

$$ln\, k_{obs} = [k_{f,H_2O} \cdot exp(-m_f \cdot [urea]) + k_{unf,H_2O} \cdot exp(m_{unf} \cdot [urea])] \qquad (4)$$

k_{obs} is the relaxation constant at a given concentration of urea, k_{f,H_2O} and k_{unf,H_2O} are the refolding and unfolding rate constants, respectively, in the absence of urea. Coefficients m_f and m_{unf} describe the urea dependence of k_{f,H_2O} and k_{unf,H_2O}, respectively.

Acknowledgments

This work was supported by the Swiss National Science Foundation and The National Center of Competence in Research "Structural Biology."

References

Batchelor, A.H., Piper, D.E., de la Brousse, F.C., McKnight, S.L., and Wolberger, C. 1998. The structure of GABP α/β: An ETS domain ankyrin repeat heterodimer bound to DNA. *Science* **279:** 1037–1041.

Binz, H.K., Stumpp, M.T., Forrer, P., Amstutz, P., and Pluckthun, A. 2003. Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332:** 489–503.

Binz, H.K., Amstutz, P., Kohl, A., Stumpp, M.T., Briand, C., Forrer, P., Grutter, M.G., and Pluckthun, A. 2004. High-affinity binders selected from designed ankyrin repeat protein libraries. *Nat. Biotechnol.* **22:** 575–582.

Bork, P. 1993. Hundreds of ankyrin-like repeats in functionally diverse proteins—Mobile modules that cross phyla horizontally. *Proteins* **17:** 363–374.

Bradley, C.M. and Barrick, D. 2002. Limits of cooperativity in a structurally modular protein: Response of the notch ankyrin domain to analogous alanine substitutions in each repeat. *J. Mol. Biol.* **324:** 373–386.

Edelhoch, H. 1967. Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochemistry* **6:** 1948–1954.

Fersht, A.R. 1999. *Structure and mechanism in protein science: A guide to enzyme and protein folding,* p. 543. W.H. Freeman and Company, New York.

Forrer, P., Stumpp, M.T., Binz, H.K., and Pluckthun, A. 2003. A novel strategy to design binding molecules harnessing the modular nature of repeat proteins. *FEBS Lett.* **539:** 2–6.

Forrer, P., Binz, H.K., Stumpp, M.T., and Pluckthun, A. 2004. Consensus design of repeat proteins. *Chem. Biochem.* **5:** 183–189.

Gomez, J., Hilser, V.J., Xie, D., and Freire, E. 1995. The heat-capacity of proteins. *Proteins* **22:** 404–412.

Groves, M.R. and Barford, D. 1999. Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.* **9:** 383–389.

John, D.M. and Weeks, K.M. 2000. Van't Hoff enthalpies without baselines. *Protein Sci.* **9:** 1416–1419.

Kohl, A., Binz, H.K., Forrer, P., Stumpp, M.T., Pluckthun, A., and Grutter, M.G. 2003. Designed to be stable: Crystal structure of a consensus ankyrin repeat protein. *Proc. Natl. Acad. Sci.* **100:** 1700–1705.

Koradi, R., Billeter, M., and Wüthrich, K. 1996. MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14:** 51–55.

Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P., and Bork, P. 2002. Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* **30:** 242–244.

Main, E.R.G., Jackson, S.E., and Regan, L. 2003. The folding and design of repeat proteins: Reaching a consensus. *Curr. Opin. Struct. Biol.* **13:** 482–489.

Makhatadze, G.I. and Privalov, P.L. 1995. Energetics of protein structure. *Adv. Protein Chem.* **47:** 307–425.

Makhatadze, G.I., Lopez, M.M., and Privalov, P.L. 1997. Heat capacities of protein functional groups. *Biophys. Chem.* **64:** 93–101.

Marcotte, E.M., Pellegrini, M., Yeates, T.O., and Eisenberg, D. 1999. A c of protein repeats. *J. Mol. Biol.* **293:** 151–160.

McDonald, N.Q. and Peters, G. 1998. Ankyrin for clues about the functi p16(INK4a). *Nat. Struct. Biol.* **5:** 85–88.

Michaely, P., Tomchick, D.R., Machius, M., and Anderson, R.G.W. Crystal structure of a 12 ANK repeat stack from human ankyrinR. *E J.* **21:** 6387–6396.

Milev, S., Gorfe, A.A., Karshikoff, A., Clubb, R.T., Bosshard, H.R., and sarov, I. 2003. Energetics of sequence-specific protein–DNA associa Conformational stability of the DNA binding domain of integrase T and its cognate DNA duplex. *Biochemistry* **42:** 3492–3502.

Mosavi, L.K. and Peng, Z.Y. 2003. Structure-based substitutions for incr solubility of a designed protein. *Protein Eng.* **16:** 739–745.

Mosavi, L.K., Minor, D.L., and Peng, Z.Y. 2002. Consensus-derived struc determinants of the ankyrin repeat motif. *Proc. Natl. Acad. Sci.* **99:** 16 16034.

Plotnikov, V.V., Brandts, J.M., Lin, L.N., and Brandts, J.F. 1997. A new sensitive scanning calorimeter. *Anal. Biochem.* **250:** 237–244.

Robinson, N.E. and Robinson, A.B. 2001. Prediction of protein deamid rates from primary and three-dimensional structure. *Proc. Natl. Acad* **98:** 4367–4372.

Sedgwick, S.G. and Smerdon, S.J. 1999. The ankyrin repeat: A diversi interactions on a common structural framework. *Trends Biochem. Sci* 311–316.

Tang, K.S., Guralnick, B.J., Wang, W.K., Fersht, A.R., and Itzhaki, L.S. Stability and folding of the tumour suppressor protein p16. *J. Mol.* **285:** 1869–1886.

Tang, K.S., Fersht, A.R., and Itzhaki, L.S. 2003. Sequential unfolding o kyrin repeats in tumor suppressor p16. *Structure* **11:** 67–73.

Walker, R.G., Willingham, A.T., and Zuker, C.S. 2000. A *Drosophila* m nosensory transduction channel. *Science* **287:** 2229–2234.

Zeeb, M., Rosner, H., Zeslawski, W., Canet, D., Holak, T.A., and Balba 2002. Protein folding and stability of human CDK inhibitor p19(INK4 *Mol. Biol.* **315:** 447–457.

Zhang, B. and Peng, Z.Y. 2000. A minimum folding unit in the ankyrin r protein p16(INK4). *J. Mol. Biol.* **299:** 1121–1132.

Zweifel, M.E. and Barrick, D. 2001a. Studies of the ankyrin repeats o *Drosophila melanogaster* Notch receptor. 1. Solution conformationa hydrodynamic properties. *Biochemistry* **40:** 14344–14356.

———. 2001b. Studies of the ankyrin repeats of the *Drosophila melanog* Notch receptor. 2. Solution stability and cooperativity of unfolding. *chemistry* **40:** 14357–14367.

Zweifel, M.E., Leahy, D.J., Hughson, F.M., and Barrick, D. 2003. Structur stability of the ankyrin domain of the *Drosophila* Notch receptor. *Pr Sci.* **12:** 2622–2632.

3 Formation of Disulfide Bonds in a Cysteine-rich Repeat Protein

In this Chapter I first review the newly detected family of cysteine-rich repeat proteins from *Helicobacter pylori* (section 3.2) followed by a general discussion on how to study disulfide bond formation in proteins (sections 3.3 and 3.4). Section 3.5 is a submitted manuscript describing my work on the oxidative refolding of the cysteine-rich repeat protein HcpB from *H. pylori.* Additional experiments on the stability and folding of HcpB are presented in section 3.6.

3.1 *Helicobacter pylori*

Helicobacter pylori (Fig. 3.1) is a micro-aerophilic, spiral shaped, gram negative bacterium discovered in 1983 by Warren and Marshall [1]. It lives in the stomach and duodenum and has a unique way of adapting to the harsh environment of the stomach. It settles in the gastric mucosa of primates and is a major risk factor for several gastric diseases such as chronic active gastritis, gastric adenocarcinoma and MALT lymphoma [2-6].

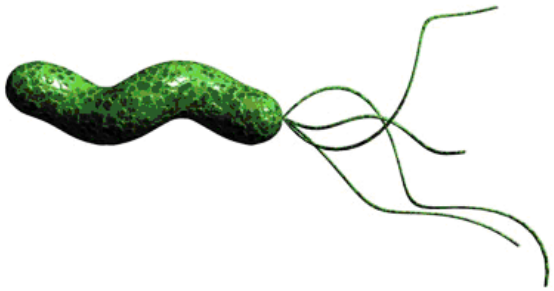


Figure 3.1. Computer-designed image of *H. pylori.* The gram negative bacterium has a curved shape and flagella that enable it to propel itself into the mucus lining of the stomach. The image was taken from http://www.faseb.org/opa/pylori/pylori.html.

Strain-specific genetic diversity has been proposed to be involved in the organism's ability to cause different diseases and to participate in lifelong chronic infection. Because of its importance as a human pathogen, the genomes of the two strains 26695 and J99 have been completely sequenced and may provide information for drug discovery and vaccine development [7, 8]. The genomes' open reading frames (ORFs) were grouped into 95 families. For approximately two-thirds of all ORFs, a function was assigned by sequence comparison. A group of seven ORFs that are unique to *H. pylori* has been identified. This group is called family 12 [7] and no clear function can yet be assigned to it. The sequences, which are annotated as "hypothetical proteins,"

may become a rich source of information if their structures and biological functions can be investigated.

3.2 Cysteine-rich repeat proteins from *Helicobacter pylori*

Sequence analysis revealed that members of family 12 possess a modular architecture of α/α-units and are rich in cysteine residues. The gene products of family 12 are designated as *Helicobacter* cysteine-rich proteins (Hcps). Hcps, which are unique to the *Helicobacter* and *Campylobacter* genera, have a mass between 15 and 40 kDa and show a conserved pattern of cysteine pairs. Two cysteine residues are separated by 7 amino acids, and there are 36 amino acids between adjacent cysteine pairs. The Hcp proteins HcpA, HcpB, HcpC and HcpE consist of, respectively, six, four, seven, and nine repeats of a common α/α-motif. Sequence conservation among Hcp proteins varies between 22 and 66% sequence identity.

Hcp proteins belong to the large group of SEL1-like tetratricopeptide repeat (TPR) proteins [9]. They seem to be involved in the Th1 cell-dependent inflammatory response exerted by *H. pylori* infection [10]. Some Hcp proteins, such as HcpA, HcpB and HcpD, bind and hydrolyse penicillin derivatives [11-13]. Whether these *in vitro* activities indicate an implication of Hcp proteins in cell-wall biosynthesis remains elusive.

3.2.1 Structures of HcpB and HcpC

The crystal structures of HcpB and HcpC (Fig. 3.2) have been solved [12, 14]. The structure of *H. pylori* cysteine-rich protein B (HcpB) consists of four α/α-motifs that are cross-linked by disulfide bridges [12]. Four disulfide bridges are observed between cysteine pairs Cys-22/Cys-30, Cys-52/Cys-60, Cys-88/Cys-96, and Cys-124/Cys-132. The disulfide bridges subdivide the structure into four pairs of α-helices named A and B. The helices have variable length. The two cysteines forming a disulfide bridge between helices A and B are located at the C-terminus of helix A and close to the N-terminus of helix B. The four α-helix pairs of HcpB have very similar conformations and the sequence identity for the pairwise alignment is 33 to 58%. Although the overall sequence composition of the first repeat is similar to that of the other repeats, the structure of the first repeat is quite different. This is due to a different loop connecting helices A and B of the first repeat.

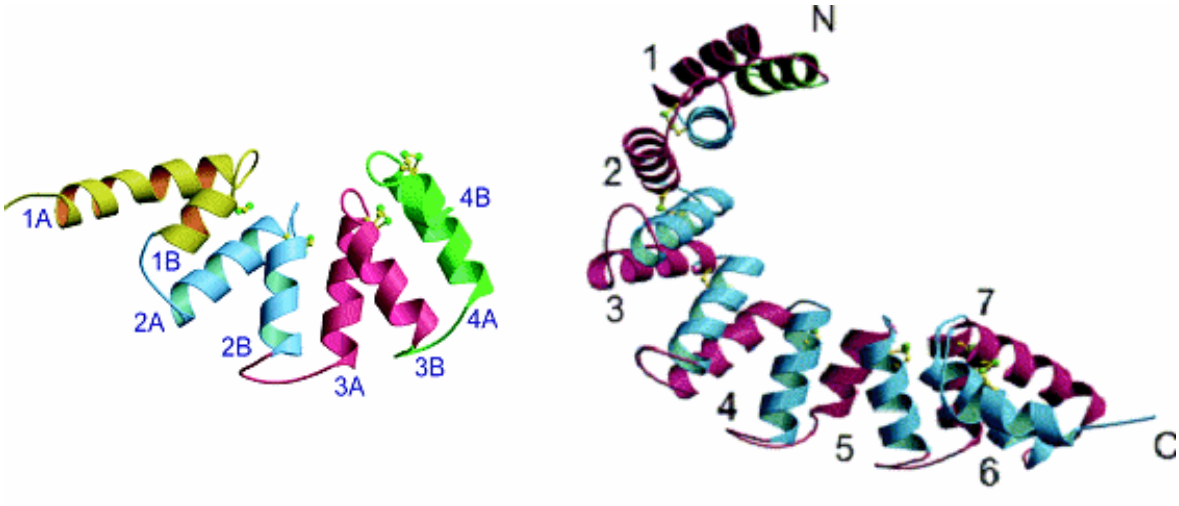


Figure 3.2. Panel A. Crystal structure of HcpB (PDB code: 1KLX). Disulfide bridges are highlighted as light green balls and the four repeats consisting of two helices each are differently coloured differently [12]. **Panel B. Crystal structure of HcpC** (PDB entry: 1OUV). The N-terminal capping helix (green), helix A (pink) and helix B (blue) of each repeat are shown. The disulfide bridges are indicated as light green balls between the helices of each repeat [14].

The crystal structure of HcpC (Fig. 3.2) shows that it consists of 15 antiparallel α-helices and is subdivided into seven repeats, each 36 residues long [14]. HcpB and HcpC share 46% sequence identity. HcpC has an N-terminal capping repeat to shield the first repeat from solvent and to stabilizes the elongated hydrophobic core. Similar N-terminal capping helices are predicted for all Hcp family members, except for HcpB, and are found frequently in solenoid proteins [15, 16]. In HcpB, the first repeat seems to act as a capping repeat.

The presence of disulfide bridges is a unique feature of Hcp proteins. Disulfide bonds have been reported only in one structurally unrelated anti-freeze repeat protein. Apart from anti-freeze repeat proteins, Hcps are the only repeat proteins that possess disulfide bridges.

3.2.2 Comparison of HcpB with tetratricopeptide repeats (TPR)

The modular architecture of Hcps shows that they belong to the large group of SEL1-like tetratricopeptide repeat (TPR) proteins. Data base searches revealed that the structure of HcpB is most similar to the TPR domain of human phosphatase 5 [17]. The isolated PP5 TPR repeats (Fig. 3.3) superimpose well onto the HcpB structure. The TPR and Hcp repeats consist of 34 and

36 amino acids, respectively. In HcpB the inter- and intra-repeat helix packing angles differ substantially from each other (14° and 42°, respectively), in contrast to the uniform packing angle of ~24° found in TPR proteins.

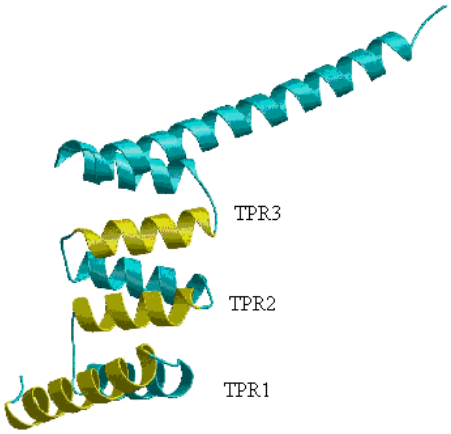


Figure 3.3. Crystal structure of tetratricopeptide repeats of human protein phosphatase 5 (PDB code: 1A17). The three repeats (TPR1, TPR2 and TPR3) and an extended C-terminal α-helix are shown [17]. The three repeats superimpose well with the repeats of HcpB.

3.3 Studying disulfide bond formation *in vitro*

Identifying intermediates and transition states on the way from the unfolded to the folded protein is a main goal of protein folding studies. For proteins containing disulfide bridges, the identification and characterization of disulfide-containing intermediates is relatively easy [18]. This is because the formation or reduction of a disulfide bond is a localized, two-state, structurally well-defined reaction and disulfide species are stable, covalent intermediates that can be isolated [19]. The rates of disulfide-bond formation and reduction can be varied without significantly altering other interactions, e.g., by changing the concentrations of the redox agents or by changing the pH. Capturing and analysing disulfide intermediates provides clues about the folding pathways of the proteins. Using disulfide bond formation to study protein folding necessitates that the folded protein conformation requires at least one disulfide for stability. In the absence of denaturant and under conditions leading to folding, the reduced protein should be at least partly unfolded [20, 21].

3.3.1 The chemistry of disulfide bond formation

Disulfide bond formation is based on thiol/disulfide exchange according to [22]:

$$R_1S^- + R_2SSR_3 \rightarrow R_2S^- + R_1SSR_3$$

The thiolate anion R_1S^- of the reducing agent displaces one sulfur of the disulfide bond R_2SSR_3 of the protein. Disulfide bonds are formed and reduced by two such thiol/disulfide exchange reactions with a redox agent, the first of which involves the formation of a mixed disulfide bond between the protein and the redox reagent. Thiol/disulfide exchange reactions can also occur intramolecularly; for example, a protein thiolate group may attack a disulfide bond of the same protein, leading to disulfide reshuffling. Redox agents like reduced and oxidized cyclic dithiothreitol ($DTT^{red} \rightleftarrows DTT^{ox}$) and oxidized and reduced linear glutathione (GSSG ⇄ 2 GSH) can be used. Disulfide bond reactions are sensitive to the concentration of the redox agents, pH of the solvent, pK_a values of the thiol groups and accessibility of the disulfides of the protein. Since thiol/disulfide exchange requires the thiolate anion, acidification quenches disulfide bond formation, disulfide reduction and thiol/disulfide exchange [23].

3.3.2 Trapping and identification of disulfide intermediates

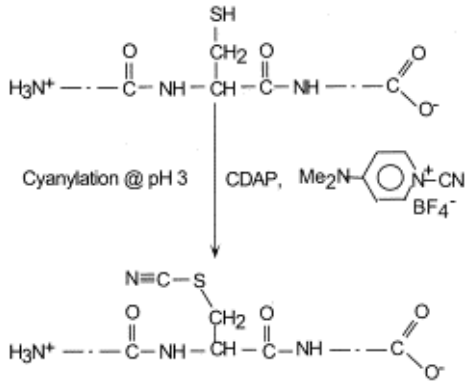
Chemical trapping – Traditionally, folding intermediates have been trapped by the addition of iodoacetate, a reagent that alkylates free thiols to prevent oxidation or thiol/disulfide exchange. Rearrangement of intermediates during trapping with iodoacetate has been observed for BPTI [23] and ribonuclease A [24]. Rearrangement occurs because thiol/disulfide exchange proceeds on the same time scale as alkylation by iodoacetate. This is a severe problem if steric hindrance retards the rate of thiol alkylation of an intermediate [25]. Although a large excess of iodoacetate can be applied to minimize the side reaction, modification of other functional groups by the high concentration of iodoacetate may cause other problems [26].

Rothwarf and Scheraga used 2-aminoethyl methanethiosulfonate (AEMTS) as a trapping agent for trapping the disulfide intermediates of bovine ribonuclease A. AEMTS reacts specifically with thiol groups and much more quickly than iodoacetate, preventing thiol/disulfide exchange [24]. Selective labeling of cysteines has also been achieved with colored and fluorescent reagents [23, 29]. To identify trapped cysteine-containing intermediates, the protein is digested by

suitable proteases, peptides are separated and analyzed by mass measurements. By such peptide mapping one can identify the fragments that contain oxidized disulfide bridges or reduced and modified cysteines.

Acid quenching – Acid quenching is an alternative technique for providing 'snapshots' of the refolding of cystinyl proteins. Because the thiolate anion is the reactive species in thiol/disulfide exchange, it is possible to rapidly quench the folding process by lowering the pH. Acid-quenched intermediates can be separated by reversed-phase high-performance liquid chromatography (HPLC) at acidic pH without significant reshuffling of the trapped species. Acid quenching is reversible. Therefore, an acid-quenched intermediate can be purified at low pH and allowed to reshuffle again at slightly alkaline pH [23, 27]. However, intermediates trapped by acid must be chemically modified to determine the disulfide bond structure.

Acid quenching and chemical trapping by cyanylation – A new method has been developed by Watson and co-workers to assign disulfide bond pairings by partial reduction, cyanylation, cleavage and mass mapping [28]. The modifying reagent is 1-cyano-4-dimethylamino-pyridinium tetrafluoroborate (CDAP). It specifically cyanylates free cysteines at acidic pH, a clear advantage over other chemical trapping reagents that usually work only at neutral to alkaline pH (Scheme 3.1). The cyanylation reaction is carried out at pH 3 to minimize disulfide exchange. Cyanylation is quantitative and specific for free sulhydryl groups. The side chains of Met, Ser, Thr, acidic and basic amino acids are not modified by CDAP. We have used acid quenching with cyanylation to trap free cysteines of HcpB during the oxidative refolding (section 3.5) and the reductive unfolding reaction (section 3.6.1).



Scheme 3.1. Cyanylation reaction of CDAP. CDAP cyanylates sulfhydryl groups at acidic pH [28].

3.4 Studying disulfide bond formation *in vivo*

In the cell, disulfide bond formation is a catalysed process [30]. In *E. coli*, disulfide bonds are introduced in the periplasm by the Dsb proteins ("Dsb" stands for *d*isulfide *b*ond formation). The oxidation of cysteine pairs of substrate proteins is achieved by the highly oxidizing DsbA/DsbB catalytic system (Fig. 3.4). The oxidizing power utilized by DsbA is delivered by DsbB, which obtains the "oxidative power" from a membrane-embedded electron transport system utilizing molecular oxygen as the terminal oxidant [31].

The DsbC/DsbD system has the ability to reduce disulfide pairs, which are then again re-oxidized by the DsbA/DsbB system (Fig. 3.5). In this way, non-native, i.e., wrongly formed disulfide bridges can be re-opened and the native disulfide pattern is found by "trial and error". The DsbC/DsbD system is sustained by a constant supply of reducing power from the cytoplasmic thioredoxin system, which utilizes NADPH as the ultimate electron source [31]

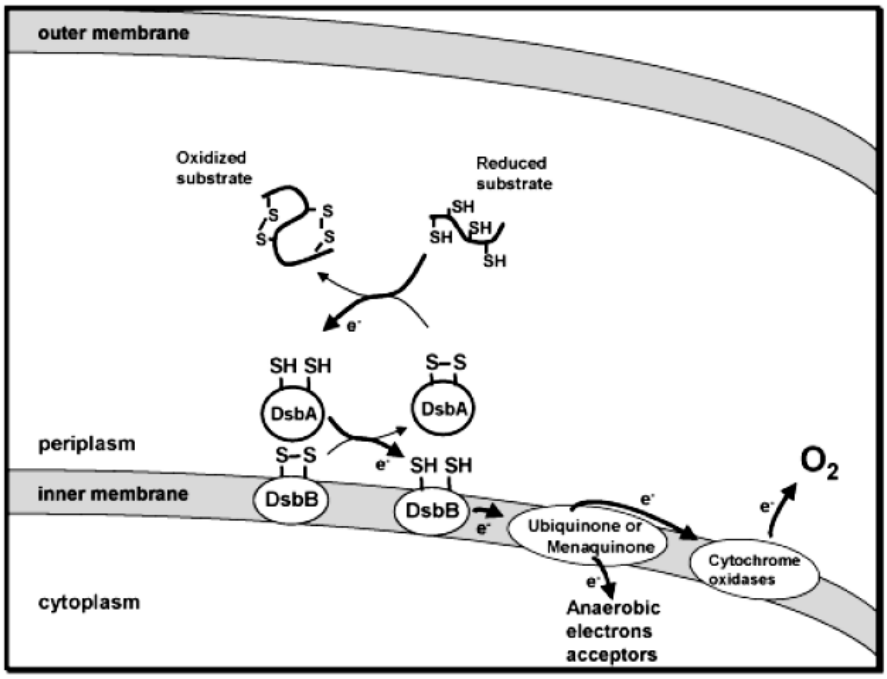


Figure 3.4. DsbA/DsbB system to oxidize cysteines in the periplasm of *E.coli*. DsbA interacts with a folding protein containing reduced cysteines, oxidizing them to form disulfide bonds. DsbA is re-oxidized by the inner membrane protein DsbB. Under aerobic conditions, DsbB is re-oxidized by ubiquinone, which passes the electrons to cytochrome oxidase and finally to molecular oxygen. Under anaerobic conditions, DsbB is re-oxidized by menaquinone, which passes the electrons to some anaerobic electron acceptor. Figure taken from [31].

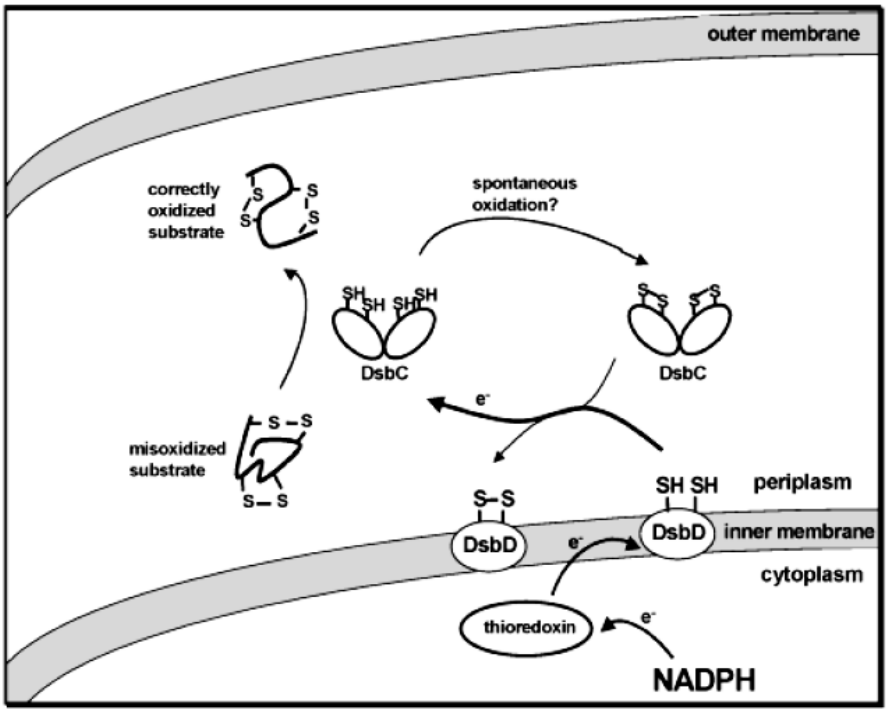


Figure 3.5. DsbD/DsbC system to reduce disulfide bonds in the periplasm of *E.coli.* DsbC reduces disulfide bonds, allowing these to rearrange to their native pairings. DsbC is reduced by DsbD. DsbD is reduced by cytoplasmic thioredoxin, which is reduced by NADPH. Figure taken from [31].

3.5 *In vitro* and *in vivo* formation of consecutive disulfide bonds in a repeat protein from *Helicobacter pylori*

V. Sathya Devi, Christine Berger Sprecher, Peter Hunziker, Peer R.E. Mittl, Patrick Frei, Hans Rudolf Bosshard and Ilian Jelesarov

(manuscript)

IN VITRO AND *IN VIVO* FORMATION OF CONSECUTIVE DISULFIDE BONDS IN A REPEAT PROTEIN FROM *HELICOBACTER PYLORI*

V. Sathya Devi¶, Christine Berger Sprecher¶, Peter Hunziker¶, Peer R. E. Mittl¶, Patrick Frei§, Hans Rudolf Bosshard¶ and Ilian Jelesarov¶

From the ¶Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, and the §Institut für Molekularbiologie und Biophysik, ETH-Hönggerberg, HPK E 17, CH-8093 Zürich, Switzerland

Helicobacter pylori **cysteine rich proteins (Hcps) are disulfide-containing repeat proteins. The repeating unit is a 36-residue helix-loop-helix motif with a disulfide bridge between the helices. We use HcpB containing four repeats arrayed in tandem as a model to study the formation of disulfide bonds in a repeat protein. *In vivo* oxidative refolding was conducted in the periplasm of *Escherichia coli*. Disulfide bonds appear consecutively in the direction of export into the periplasm and there is no evidence of rearrangement of non-native disulfide bonds. This follows from the observation that folding needs only DsbA and DsbB, the catalysts of disulfide bond formation, but not DsbC and DsbD, the catalysts for rearranging non-native disulfide bonds. *In vitro*, the oxidative refolding in the presence of glutathione redox buffer also proceeds without the apparent formation of incorrect disulfide bonds. Three intermediates were observed: $HcpB_{1SS}$, $HcpB_{2SS}$ and $HcpB_{3SS}$ with one, two and three disulfide bonds, respectively. Intermediates $HcpB_{1SS}$ and $HcpB_{2SS}$ are molecular mixtures with correct disulfide bonds in any one of the four repeats. $HcpB_{3SS}$ is the major folding intermediate with disulfides in repeats 2, 3 and 4 and reduced cysteines in repeat 1. Rate-limiting formation of the last disulfide bond is strongly coupled to the folding of the N-terminal repeat 1 and contributes about half of the thermodynamic stability of HcpB, which is 27 kJ mol^{-1} at 25 °C and pH 7. This may be explained by the N-terminal repeat acting as a cap to stabilize the folded molecule.**

Repeat proteins are built of structurally identical motifs arranged in tandem to form elongated shapes as found, for example, in ankyrin, tetratricopeptide or leucine rich repeat proteins. Such proteins are involved in signaling and regulatory pathways and function as versatile scaffolds for protein-protein interactions (1). Designed repeat proteins have great potential as binding partners (2,3). The repeating motifs typically have 20–40 residues forming secondary structures that coalesce in various topologies (4). The linear stacking of repeat modules produces local, short-range packing interactions at the inter-repeat interface.

A family of cysteine containing repeat proteins has been detected as open reading frames in the genome of *Helicobacter pylori* (5,6). Gram negative *H. pylori* has been made responsible for several gastric and duodenal diseases (7,8). The precise biological function of Hcps [1] is not known, but it was found that some Hcps bind β-lactam compounds (6,9) and seem to be involved in the innate immune response (10). The proteins are built of several consecutive, 36-residue long α/α motifs, the anti-parallel α-helices of the motif being connected by a short loop. The structure belongs to the Sel1-like protein family, a subfamily of the tetratricopeptide repeat proteins (1). The α-helices of the Hcp motif are bridged by a disulfide bond, a unique feature not found in

other repeat proteins. The shortest Hcp family member is the 16 kDa protein HcpB with four repeat units and four disulfides bridges (Fig. 1) (11).

The formation of an array of motifs of a repeat protein may differ from the folding of a similarly sized globular protein because molecular packing in a repeat protein occurs mainly by consecutive short-range interactions. Intra-repeat and inter-repeat interactions affect the folding mechanism and the overall stability increases with the number of repeats (12,13). There are few detailed studies on the stability and folding kinetics of repeat proteins. Even in the simple case of an ankyrin with only three repeats, rapid folding seems to proceed through intermediates (14). The folding of designed tetratricopeptide repeat proteins was shown to be governed by interactions within and between the repeating motifs and is very rapid (13).

Protein disulfide formation *in vivo* has been well characterized (short reviews in (15,16)). In prokaryotes, an oxidative pathway is responsible for the formation of disulfides and an isomerization pathway for the reshuffling of incorrectly formed disulfides (15). DsbA is the protein responsible for the oxidation of cysteines to cystine in the periplasm. The strongly oxidizing enzyme forms disulfides between any two cysteines that happen to be close enough to each other. This was demonstrated *in vitro* for ribonuclease A, which contains four non-sequential disulfide bonds. DsbA (in the presence of DsbB and quinones to provide oxidizing equivalents) fully oxidizes ribonuclease A yet the oxidized enzyme is inactive as it contains incorrect disulfide bonds and is misfolded (17). Hence, there is a need for a reductive isomerization pathway in the periplasm. Its main player is disulfide bond isomerase DsbC. This enzyme rapidly reduces disulfides to give the folding protein the opportunity for disulfide reshuffling, until the correct structure has been reached. This double mechanism of oxidation and reshuffling has been confirmed: Folding of proteins with non-sequential disulfide bonds needs both DsbA and DsbC (plus DsbB, DsbD, thioredoxin, quinones, etc. to provide oxidizing and reducing equivalents to DsbA and DsbC, respectively). On the other hand, proteins with sequential disulfide bonds are correctly formed in the absence of disulfide isomerase DsbC. In other words, there is no need for reshuffling of wrong disulfides when the cysteines appear in the correct order as the protein is exported to the periplasm (18).

Studies on *in vitro* oxidative refolding of protein disulfides have a long history. Already in Anfinsen's seminal paper of 1961 it was conjectured that oxidative refolding is a dynamic process involving rapid closing and opening of disulfide bonds to reach the correct pairings by "trial and error" (19). Subsequent work on ribonuclease, bovine pancreatic trypsin inhibitor, lysozyme and other disulfide-containing proteins amply demonstrated the dynamic nature of disulfide bond formation and the temporary rise and fall of incorrect disulfides (recent review in (20)). Hence, efficient oxidative refolding in the test tube needs a redox buffer, for example, reduced and oxidized glutathione, to speed up disulfide formation and enable disulfide reshuffling.

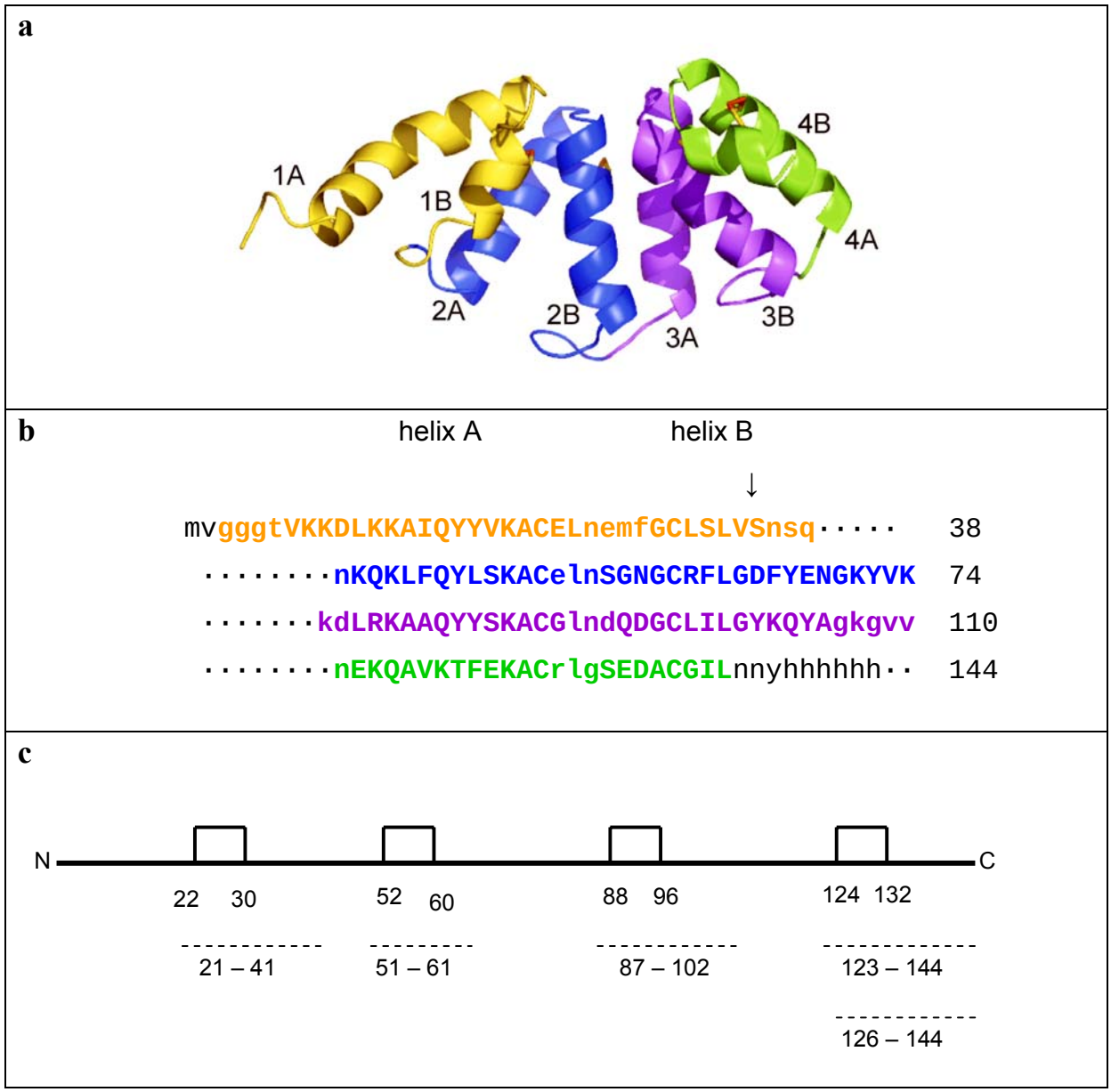


FIG. 1. **Structure of HcpB.** (a) Crystal structure of HcpB. The four repeat motifs are colored differently and the disulfide bridges are shown in red. Note that the assignment of helices A and B to repeats 1 to 4 follows the definition of Luthy et al. (11), which differs from the assignment used for tetratricopeptide repeats (1). (b) Sequence of HcpB with C-terminal His-tag. The four repeats are colored as in panel (a) and aligned for sequence homology. Residues in lower case are not part of helices, black residues are not seen in the 3D-structure shown in panel (a). The arrow indicates cleavage after Val^{34} during synthesis of pDsbA-HcpB in *E. coli* (see the text for details). (c) Pattern of sequential disulfide bonds (boxes with position of cysteines indicated) and of the cysteine-containing tryptic peptides (dashed lines). Numbers indicate the N- and C-terminal residues of the peptides. Peptide 126–144 is produced only if Cys^{124} is reduced, peptide 123–144 only if disulfide Cys^{124}–Cys^{132} is present.

The contribution of disulfide bonds to the stability of Hcps and the mechanism of disulfide formation in Hcps are unknown. We have chosen HcpB, which includes four disulfides, to study the oxidative folding of a repeat protein both *in vivo* and *in vitro*. *In vivo* folding of HcpB in the periplasm of *E. coli* follows the direction of export into the periplasm and needs no enzymes for disulfide rearrangement. In the test tube one may envisage two principle modes of folding. If the disulfides contribute very significantly to the stability of the repeat motif, that is, if the reduced protein is essentially unstructured, incorrect disulfides are expected to arise. If, however, the reduced protein exhibits some repeating structural features before disulfides are formed, incorrect disulfides should have only a small chance of building up. This we could confirm: The first two correct disulfides form at random, followed by the appearance of a major, structurally unique intermediate with three disulfide bonds. Final formation of fully oxidized, native HcpB is rate-limited by disulfide formation in the N-terminal repeat.

EXPERIMENTAL PROCEDURES

Buffers and chemicals – Standard buffer: 20 mM sodium phosphate, 100 mM KCl, pH 7.0. Redox buffer: standard buffer containing 2 mM GSH and 1 mM GSSG. Periplasmic extraction buffer: 50 mM sodium phosphate, 300 mM NaCl, pH 8.0, with 1 mg mL^{-1} polymyxin B (Sigma). Chemicals were of analytical grade or of the best grade available and were used without further purification.

Expression and purification of HcpB – Recombinant HcpB was expressed in *E. coli*, solubilized from inclusion bodies and affinity-purified on Ni-NTA agarose as described (11). The protein was desalted and freeze dried and the mass of 16160.5 Da, including the 6-His tag (Fig. 1b), was confirmed by mass spectrometry. The concentration was determined by UV spectroscopy using $\varepsilon_{276} = 15080\ cm^{-1}\ M^{-1}$.

Cloning of DsbA-HcpB – To study HcpB folding in the periplasm of *E. coli*, the gene sequence of HcpB was N-terminally fused with the signal sequence of DsbA. Plasmid pDsbA3 (21) harbors the DsbA gene under the control of the trc promoter. A unique *Nhe*I restriction site is located at the end of the DsbA signal sequence and a unique *Bam*HI restriction site at the end of the DsbA gene. *Nhe*I and *Bam*HI restriction sites were introduced by PCR up- and downstream of the HcpB gene using the plasmid pTFT74/HP0336 (11), the forward primer 5'-AGCGTTTAGCGCTAGC-GCGATGGTAGGGGGTGGAACGGTA-3' (*Nhe*I site underlined) and the reverse primer 5'-CGGGATCCTTAGTGGTGGTGGTGGTGGTGGTAG-3' (*Bam*HI site underlined). The resulting PCR fragment of the HcpB gene and the pDsbA3 plasmid were each digested with *Nhe*I and *Bam*HI. The pDsbA3 plasmid was dephosphorylated with shrimp alkaline phosphatase (Roche Diagnostics) to avoid self-ligation. After purification (Qiagen kit) the HcpB PCR fragment and the pDsbA3 plasmid were ligated with T4 DNA ligase for 1 h at room temperature. *E. coli* strain XL1-Blue chemically competent cells were transformed with the ligation mixture and colonies were grown overnight on LB medium supplemented with 50 μg mL^{-1} ampicillin. Plasmid DNA of individual colonies was isolated by the Qiagen mini prep kit from saturated overnight cultures. The final plasmid pDsbA-HcpB features the trc promoter, the DsbA signal sequence, the HcpB gene and a C-terminal 5×His-tag. Its sequence was confirmed by DNA sequencing.

In vivo folding of HcpB – Cells of *E.coli* MC 1000 (wild type), and mutant strains JCB817 (MC1000, lambda 102+ $DsbA^-$), JCB819 (MC1000, lambda 102+ $DsbB^-$), FED 126 (MC1000, $DsbD^-$) and FED 215 (MC1000, DsbC::mini-Tn10KanR) were transformed with the pDsbA-

HcpB plasmid and grown overnight at 37 °C. A single colony was picked from each plate and grown in 5 mL of 2YT medium overnight (30 °C, 180 rpm agitation). Main cultures (500 mL) were grown under the same conditions. When an OD of 0.6 was reached, expression was induced by 1 mM IPTG, and the culture was grown for an additional 2 or 6 h. Equal amounts of cells (concentrations adjusted by OD measurement) were centrifuged at 4 °C and 5000 rpm for 20 min. Pellets were re-suspended in 10 mL of periplasmic extraction buffer, shaken for 1 h at 4 °C and centrifuged at 18000 rpm for 30 min. Periplasmic extracts were mixed with sample buffer, boiled for 5 minutes at 95 °C and separated by electrophoresis on 15% SDS-PAGE. Western blots were prepared with monoclonal mouse anti-tetra-His IgG1 (primary antibody, Qiagen) and goat anti-mouse IgG conjugated with horse radish peroxidase (secondary antibody, Jackson Immuno Research Laboratories, Inc.). Bands were visualized by chemiluminescence (BM chemiluminescence POD, Roche Applied Science).

In vitro oxidative refolding – HPLC-purified, recombinant HcpB (1 mg mL^{-1}) was treated with 5 M GdmCl and 0.1 M DTT in standard buffer for 2 h at room temperature. GdmCl and DTT were rapidly removed by passing the reaction mixture through a Hi-trap desalting column (Sephadex G-25, Amersham Biosciences), which had been pre-equilibrated in standard buffer. The eluate was captured in freshly prepared redox buffer to start oxidative refolding. The final concentration of HcpB in the redox buffer was 0.5 mg mL^{-1}. Samples were removed at timed intervals and cysteine oxidation was quenched by adding an equal volume of 8% TFA. Samples were chromatographed by RP-HPLC on a Vydac C8 analytical column (4.6 mm × 250 mm, Nucleosil 300-5). The column was eluted with a linear gradient from 46% to 54% B in A at a flow rate of 0.7 mL min^{-1}. (A is 3% acetonitrile, 0.1% TFA in water; B is 80% acetonitrile, 0.085% TFA in water.) Peaks were collected, lyophilized and stored at –20 °C for further analysis.

Trapping of intermediates by cyanylation – Lyophilized fractions from HPLC-separation (0.5 mg mL^{-1}) were cyanylated with 0.2 M freshly prepared CDAP in 100 mM sodium citrate buffer, pH 3.0, for 20-30 min at room temperature. CDAP was present in at least 20 fold molar excess over free cysteines. Cyanylated material was immediately re-chromatographed by RP-HPLC to remove the reagent and to re-purify the cyanylated protein fractions. Cyanylation did not markedly change the retention time in RP-HPLC.

Tryptic digestion – Cyanylated HcpB was digested with trypsin (trypsin/HcpB ratio 1:100) in 20 mM Tris-HCl, 50 mM NaCl, pH 7.5 for 16 hours at room temperature. The digestion mixture was mass-analyzed by MALDI-TOF.

Major intermediate $HcpB_{3SS}$ and truncated $HcpB_{\Delta 1}$ – $HcpB_{3SS}$ purified by RP-HPLC was kept in argon-saturated standard buffer for use in CD spectroscopy and urea unfolding studies. The concentration was determined from ε_{276} = 15080 cm^{-1} M^{-1}. $HcpB_{\Delta 1}$ was prepared from *in vivo* synthesis. To this end, *E.coli* MC 1000 cells that had been transformed with pDsbA-HcpB were grown overnight at 30 °C after induction with 1 mM IPTG. The periplasmic extract was prepared as described above and purified by Ni-NTA chromatography followed by RP-HPLC fractionation. The peak corresponding to $HcpB_{\Delta 1}$ was identified by ESI-MS and automated Edman degradation, and the concentration was calculated from ε_{276} = 12035 cm^{-1} M^{-1}.

Urea unfolding – Unfolding by urea was followed by the change of the CD signal at 222 nm. Spectra were taken after overnight incubation of the protein at the desired urea concentration. In the case of intermediate $HcpB_{3SS}$, cysteine oxidation was prevented by keeping all solutions saturated with argon. The integrity of $HcpB_{3SS}$ after urea unfolding was confirmed by RP-HPLC and ESI-MS. Urea unfolding data were analyzed according to $[\theta] = (K_U/(1 + K_U))([\theta_U] - [\theta_F]) + [\theta_F]$, where K_U is the unfolding constant defined as $f_U/(1 - f_U)$ (f_U = fraction of unfolded protein),

[θ] is the molar ellipticity per residue at 222 nm, and [θ_F] and [θ_U] are the molar ellipticity per residue of folded and unfolded protein, respectively, assumed to be linear functions of [urea] of the form [θ_i] = [$\theta_{i,0}$] + α_i[urea]. The free energy of unfolding in the absence of urea, ΔG_U^W, was extrapolated from $\Delta G_U = \Delta G_U^W - m$[urea], where $\Delta G_U = -RT\ln K_U$.

Instrumentation – ESI-MS was performed on a QTOF Ultima API (Waters) instrument in 0.2% formic acid, 50% acetonitrile at a flow rate of 0.8 μL min^{-1}. MALDI-TOF analysis was performed on a Biflex III (Bruker) instrument, using the α-cyano-4-hydroxy cinnamic acid matrix. CD spectra were measured on a J-715 spectropolarimeter (Jasco) equipped with a temperature-controlled water bath.

RESULTS

In vivo folding of HcpB produces no wrong disulfides – Folding was studied in the periplasm of *E. coli*. The HP0336 open reading frame for HcpB in the genome sequence of *H. pylori* strain 26695 has no leader sequence for export into the periplasm (22). Therefore, HcpB was N-terminally fused with the export signal sequence of DsbA. The pDsbA-HcpB construct, which also contained a C-terminal His-tag, was well expressed in wild type *E. coli* strain MC1000 (Fig. 2). The strain has the full set of enzymes necessary to oxidize cysteines (DsbA and DsbB) and to isomerize disulfides (DsbC and DsbD). HcpB is expressed to a similar degree in the mutant strains lacking either disulfide isomerase DsbC or the partner enzyme DsbD that keeps DsbC reduced. In contrast, no HcpB is found in mutant cells lacking either the oxidizing enzyme DsbA or the partner enzyme DsbB that keeps DsbA oxidized. This finding demonstrates very clearly that oxidative folding of HcpB in the periplasm of *E. coli* does not pass through incorrect disulfides before reaching its native structure.

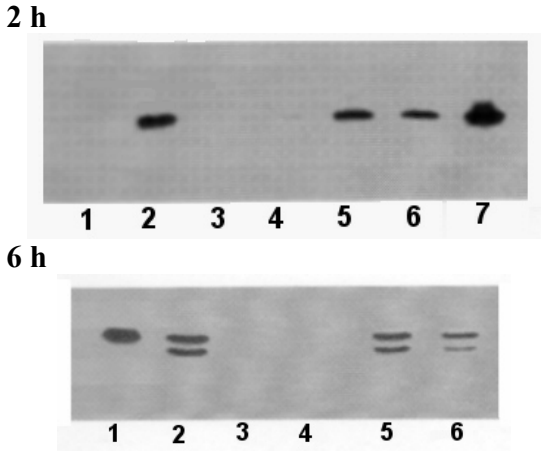


FIG. 2. **Western blots of periplasmic extracts from the synthesis of HcpB in *E. coli*.** (Top) Extracts from cells induced by IPTG for 2 h. Lane 1, negative control (exported Semliki forest virus protein containing no disulfide and no His-tag); lane 2, wild type; lane 3, $DsbA^-$; lane 4, $DsbB^-$; lane 5, $DsbC^-$; lane 6, $DsbD^-$; lane 7, positive control (purified HcpB, overloaded). (Bottom) Extracts from cells induced by IPTG for 6 h. Lane 1, positive control; lane 2, wild type; lane 3, $DsbA^-$; lane 4, $DsbB^-$; lane 5, $DsbC^-$; lane 6, $DsbD^-$.

When cells were induced for 6 h, a second, faster moving band appeared (Fig. 2 lower panel). Both bands were purified by Ni-NTA affinity chromatography. The mass of the upper band and its N-terminal sequence correspond to full length HcpB with a mass[2] of 16023 kDa and an N-terminal sequence of MVGGGT. The lower band has a mass of 12305 kDa and starts with

the sequence SNSQIN. Thus, the second band is a truncated form of HcpB lacking residues 1–34. It is named $HcpB_{\Delta 1}$ to indicate the missing first repeat (Fig. 1). Truncation most likely resulted from proteolysis in the periplasm. In support, essentially the same truncated product is produced when native HcpB is digested with thermolysine for up to 1 h at pH 7 and 25 °C. Thermolysine cleaves the native protein exclusively after Ser^{32} and leaves a folded intact protein with the sequence 33–144, which is virtually identical to $HcpB_{\Delta 1}$ formed under prolonged *in vivo* folding in *E. coli* (not shown).

In vitro oxidative refolding of HcpB follows a sequential pathway with a dominant three-disulfide intermediate – In principle, disulfide formation can be detected by rapidly and irreversibly blocking and identifying the free thiol groups remaining at different time points after the start of the folding reaction. Acidification is another means of quenching disulfide bond formation since disulfide formation occurs through the deprotonated thiol group. We have combined acidification with chemical blocking to follow the *in vitro* folding pathway of HcpB. Sulfhydryl groups were blocked by cyanylation with 1-cyano-4-dimethylamino-pyridiniumtetrafluoroborate (CDAP) (23). Summarized briefly, the procedure was as follows: Fully reduced and chemically denatured HcpB was rapidly transferred into glutathione redox buffer at pH 7.0. Refolding was allowed to proceed at ambient temperature. Samples were taken at timed intervals, disulfide formation was quenched with TFA and samples were analyzed by RP-HPLC under acidic conditions to prevent further oxidation. Chromatographic peaks were collected and cyanylated with CDAP at pH 3 to block sulfhydryl groups. Finally, the cyanylated products were digested with trypsin and the digestion products identified by MALDI-TOF.

Oxidative refolding was started by a rapid buffer change (~15 s) on a short desalting column. Folding started immediately as indicated by the rapid drop of the CD signal at 222 nm. About half of the native CD signal at 222 nm was reached after 30 s (not shown). Appearance of oxidized protein took considerably longer. Fig. 3 shows the time course of disappearance of reduced HcpB and appearance of oxidized HcpB during refolding in glutathione buffer. Table 1 shows the mass analysis of the chromatographic peaks of Fig. 3. Reduced protein (R) dominates during the first few minutes. Minor intermediates $HcpB_{1SS}$ and $HcpB_{2SS}$ appear shortly after the start of the folding reaction and are gone when the major intermediate $HcpB_{3SS}$ peaks. Oxidized protein (N) starts to appear after about 1 h. Mass analysis shows that $HcpB_{1SS}$, $HcpB_{2SS}$ and $HcpB_{3SS}$ contain 1, 2 and 3 disulfides, respectively (Table I). To identify the location of the disulfide bonds of the intermediates, the peaks labeled $HcpB_{1SS}$, $HcpB_{2SS}$, $HcpB_{3SS}$, R and N were cyanylated and digested with trypsin. Tryptic peptides with cysteine residues were identified by MALDI-TOF to obtain the results shown in Table II. Peak N contains only peptides with intact disulfide bonds, that is, no cyanylated thiol groups. Peak R contains all the cyanylated peptides, that is, no disulfides. Peaks $HcpB_{1SS}$ and $HcpB_{2SS}$ are mixtures of all the cyanylated and non-cyanylated tryptic peptides. This means that, at the start of the oxidative folding reaction, a correct disulfide can form in any of the four repeats. In contrast, peak $HcpB_{3SS}$ is a unique species. Its digestion by trypsin produces oxidized (non-cyanylated) peptides from repeats 2, 3 and 4, and one reduced (cyanylated) peptide from the first repeat. Most important, no other cyanylated or non-cyanylated peptides were found. Incidentally, the non-cysteine containing tryptic peptides could also be identified by mass analysis (not shown).

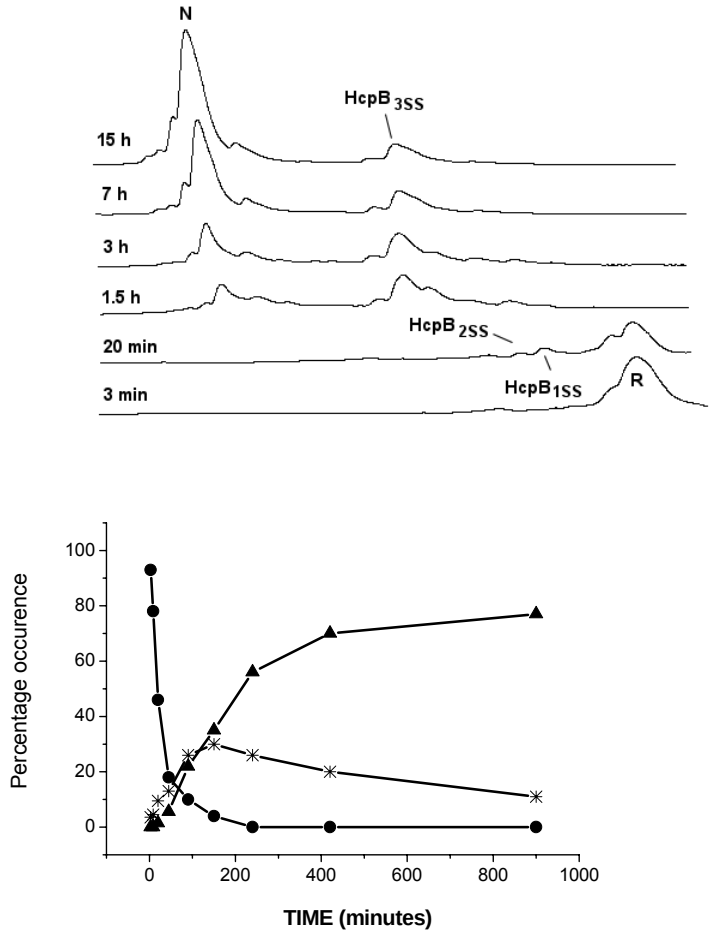


FIG. 3. **Oxidative refolding of HcpB in glutathione buffer at pH 7.0 and room temperature.** (Top) Separation of reaction products by RP-HPLC after different times of refolding. R and N are the fully reduced and fully oxidized proteins, respectively. Minor peaks to the left and right of N have the same mass as N; the minor peak preceding $HcpB_{3SS}$ has the same mass as $HcpB_{3SS}$; the shoulder preceding R has the same mass as R; the minor peak following $HcpB_{3SS}$ in the 1.5 h and 3 h traces is a molecular mixture that could not be resolved. (Bottom) Disappearance of reduced protein R (●), appearance of the major intermediate $HcpB_{3SS}$ (*) and of the oxidized protein N (▲).

Fully reduced HcpB is partially folded – If the fully reduced protein had a random structure, one would expect non-native disulfide bonds to transiently arise during oxidative refolding. As this was not seen, the fully reduced protein may already be partially structured. This is confirmed by the CD spectrum shown in Fig. 4. The helix content of HcpB measured in the presence of 15 mM DTT is about 50% of that of native HcpB. Thus, the reduced protein contains significant regular structure. The main folding intermediate, $HcpB_{3SS}$, has about 85–90% of the helix content of native HcpB (dotted spectrum in Fig. 4). This agrees with its presumed structure: repeats 2, 3 and 4 folded and oxidized, repeat 1 reduced and unfolded. The three repeats of the truncated protein $HcpB_{\Delta 1}$ are likely to be folded as in the native protein since $HcpB_{\Delta 1}$ and native HcpB exhibit strongly overlapping spectra (Fig. 4).

TABLE I

Mass (Da) of reaction products of oxidative refolding separated by RP-HPLC as shown in Fig. 3

peak in Fig. 3	mass [a] (calculated)	cyanylated thiol groups / disulfides
R	16368.2 (16368.6)	8/0
$HcpB_{1SS}$	16316.3 (16316.6)	6/1
$HcpB_{2SS}$	16264.3 (16264.6)	4/2
$HcpB_{3SS}$	16212.0 (16212.5)	2/3
N	16160.5 (16160.5)	0/4

[a] Masses shown are the average masses calculated by the program MaxEnt1 (Micromass U.K.). The mass difference between a disulfide (cystine) and two cyanylated cysteines is 52.

The first repeat contributes most to the stability of HcpB – Values of ΔG_U^W, the free energies of unfolding in standard buffer at pH 7 and 20 °C, were estimated from reversible unfolding by urea. The results are shown in Fig. 5. In terms of the midpoint urea concentration of unfolding ($[urea]_{50}$, see the legend to Fig. 5), native HcpB is the most stable species, fully reduced HcpB is the least stable, and the stabilities of $HcpB_{\Delta 1}$ and $HcpB_{3SS}$ are in between. Free energies of unfolding were evaluated by the linear extrapolation method (24) assuming a simple two-state folding model. This was justified for native HcpB and truncated $HcpB_{\Delta 1}$, which unfold cooperatively (m-values ~7 kJ mol^{-1} M^{-1}). Unfolding of the major intermediate $HcpB_{3SS}$ was less cooperative (m-value 3.8 kJ mol^{-1} M^{-1}) and may pass through more than a single transition. Therefore, the value of ΔG_U^W from linear extrapolation has to be considered as a rough approximation. The unfolding curve of reduced HcpB was not evaluated since it showed no plateau at low urea concentration.

The major intermediate $HcpB_{3SS}$ and the truncated protein $HcpB_{\Delta 1}$ are about half as stable as native HcpB: ΔG_U^W = 11 – 14 kJ mol^{-1} for $HcpB_{3SS}$ and $HcpB_{\Delta 1}$ versus ΔG_U^W = 27 kJ mol^{-1} for intact HcpB. When compared to fully reduced HcpB, the molar ellipticity of the folding intermediate $HcpB_{3SS}$ amounts to about ¾ of the molar ellipticity of native HcpB (Fig. 4). Hence, reduced repeat 1 of $HcpB_{3SS}$ is probably unstructured and it is the rate-limiting folding of the first repeat that contributes very strongly to the overall stability of native HcpB. We also note that in the absence of oxygen, the major intermediate with its unfolded repeat 1 is stable for days (argon-saturated standard buffer, pH 7, 25 °C). Therefore, the folding of repeat 1 is strongly coupled to

cysteine oxidation. Put differently, it is not the folding per se but the formation of the disulfide bridge of repeat 1 that contributes very strongly to the stability of the native protein.

TABLE II

Mass[a] (Da) of cysteine containing peptides obtained by tryptic digestion of peaks in the RP-HPLC profile of oxidative refolding (Fig. 3)

peptide [cyteines] (calculated mass)	N[b]	R[b]	$HcpB_{1SS}$[b]	$HcpB_{2SS}$[b]	$HcpB_{3SS}$[b]
21–41 [C^{22}, C^{30}] (2298.1)	2297.9 (0)	2349.8 (2)	2297.6 (0) 2349.7 (2)	2297.5 (0) 2349.7 (2)	2349.8 (2)
51–61 [C^{52}, C^{60}] (1121.5)	1121.5 (0)	1173.4 (2)	1121.2 (0) 1172.3 (2)	1121.1 (0) 1172.2 (2)	1121.4 (0)
87-102 [C^{88}, C^{96}] (1680.8)	1680.8 (0)	1732.7 (2)	1680.5 (0) 1732.5 (2)	1680.4 (0) 1732.4 (2)	1680.6 (0)
123–144 [C^{124}, C^{132}] (2537.1)	2537.0 (0)		2536.7 (0)	2536.7 (0)	2536.7 (0)
126–144[c] [C^{132}] (2191.0)		2215.8 (1)	2215.6 (1)	2215.4 (1)	

[a] Monoisotopic masses of singly charged peptides, i.e. $[M + H]^+$; the mass difference between a disulfide (cystine) and two cyanylated cysteines is 52. [b] Number of cyanylated thiol groups in parenthesis; [c] The C-terminal fourth repeat yielded two cysteine-containing tryptic peptides; peptide 123–144 was produced only when the disulfide Cys^{124}–Cys^{132} was intact and peptides 123–125 and 126–144 only when the disulfide was reduced; peptide 123–125 was not determined.

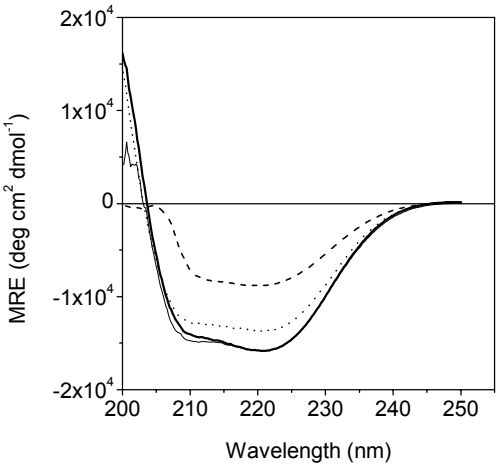


FIG. 4. **CD spectra of HcpB and its derivatives measured at pH 7 and 20 °C.** Thick solid line: native HcpB. Thin solid line: $HcpB_{\Delta 1}$ lacking the first repeat motif. Dotted line: major folding intermediate $HcpB_{3SS}$. Dashed line: fully reduced HcpB in the presence of 15 mM DTT.

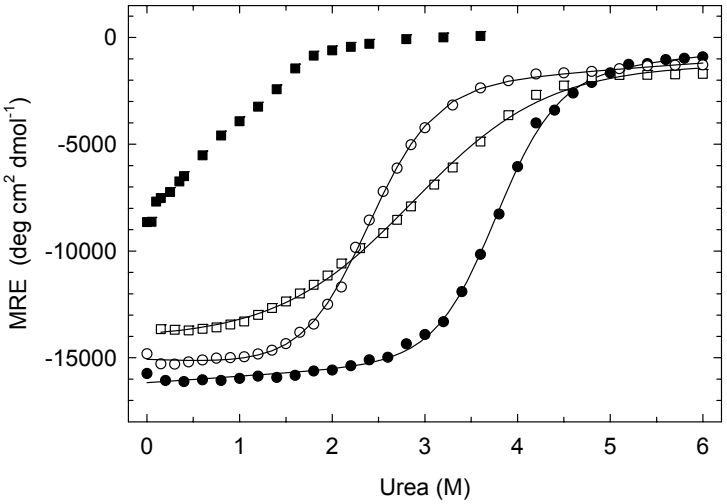


FIG. 5. **Urea unfolding curves measured at pH 7 and 25 °C.** Solid lines are best fits for two-state unfolding for: native HcpB (●), ΔG_U^W = 27.3 kJ mol^{-1}, m = 7.3 kJ mol^{-1} M; truncated $HcpB_{\Delta 1}$ (○), ΔG_U^W = 14.0 kJ mol^{-1}, m = 6.1 kJ mol^{-1} M^{-1}; major intermediate $HcpB_{3SS}$ (□), ΔG_U^W = 11.3 kJ mol^{-1}, m = 3.8 kJ mol^{-1} M^{-1}. Errors of ΔG_U^W are estimated at ±3 kJ mol^{-1}. Unfolding of fully reduced HcpB (■) at 5 °C could not be analyzed as a two-state transition.

DISCUSSION

Folding of HcpB in the periplasm of E. coli – Two enzyme systems are necessary for the correct oxidative folding of disulfide-containing proteins in the prokaryotic cell (15). The DsbA/DsbB system oxidizes cysteines as they appear in the periplasm. If disulfide bonds are not consecutive in the amino acid sequence, non-native or incorrect disulfide bonds may form by the action of DsbA/DsbB. To produce the native disulfide pattern, the DsbC/DsbD system reduces disulfide bonds for re-oxidation by DsbA/DsbB. Direct experimental support of the different tasks assigned to the oxidizing and reshuffling enzyme systems was published very recently. Correct folding of non-consecutive disulfide bonds in the *E. coli* periplasmic protein phytase needs both enzymatic systems, while mutant phytase with consecutive disulfide bonds folds in the presence of the DsbA/DsbB system alone (18). This finding of Beckwith and coworkers is now confirmed in a most straightforward way by our results: The formation of disulfide bonds located in four repeat motifs arranged in tandem needs no disulfide reshuffling and depends solely on the DsbA/DsbB system. To obtain this result, HcpB had to be equipped with a leader sequence to be secreted out of the cell. Incidentally, all other Hcps except HcpB carry an N-terminal export signal, but some Hcps, for example HcpA, were found to also be expressed in the cytoplasm despite containing an export signal (10).

Oxidative refolding of HcpB in the test tube – A vast body of work on *in vitro* folding of protein disulfides has accumulated over the years (20). Typically, non-native disulfides are found in many refolding pathways and efficient folding in the test tube needs a redox catalyst to help reshuffling of wrongly formed disulfides. Some folding pathways seem to depend on reaction conditions, and occasionally even on how the folding intermediates are being trapped (20). One feature is rather common: formation of the last disulfide bond often is rate-limiting, the conformation of the rate-limiting intermediate being close to the native state (20).

The oxidative refolding of HcpB is a rather simple process. First, intermediates appear with correct disulfide bonds in one or two of any of the four repeats. Thereafter, the unique three-disulfide species $HcpB_{3SS}$ builds up as the major folding intermediate to finally be transformed into the native protein in a rate-limiting step. Obviously, the last oxidation reaction $HcpB_{3SS} \rightarrow$ HcpB has a high transition energy barrier. Based on its CD spectrum, the first repeat of the major intermediate $HcpB_{3SS}$ is essentially unstructured. Moreover, the disulfide bond of the first repeat is completely buried in the folded protein. Taken together, this may explain why folding of the first repeat has a particularly high energy barrier.

But why should the first repeat contribute almost as much to stability as repeats 2, 3 and 4 together? There is a peculiarity to HcpB. Most Hcps consist of an odd number of helices where the first N-terminal helix acts as a capping structure to protect the hydrophobic core of the protein from the solvent. Not so in HcpB. Here the first repeat by itself acts as a capping motif. This may explain the higher stability contribution of the first repeat. As a capping structure it protects the core of the molecule from the aqueous environment, thereby adding much to stability.

There is another irregularity to repeat 1. The geometry of the contact between repeats 1 and 2 differs significantly from the other inter-repeat contacts, as shown in Fig. 6. In the figure, interactions between helix B of one repeat and helix A of the following repeat (indicated as helices A* in Fig. 6) are being compared. Interactions between repeats 2 and 3 and between repeats 3 and 4 are strong and super-imposable. The corresponding interaction between helix B of repeat 1 and helix A* of repeat 2 differs significantly. An altered geometry between the first two repeats may have to do with the capping function of the first repeat and may also explain the limited proteolysis of a single peptide bond in helix 1B producing truncated $HcpB_{\Delta 1}$.

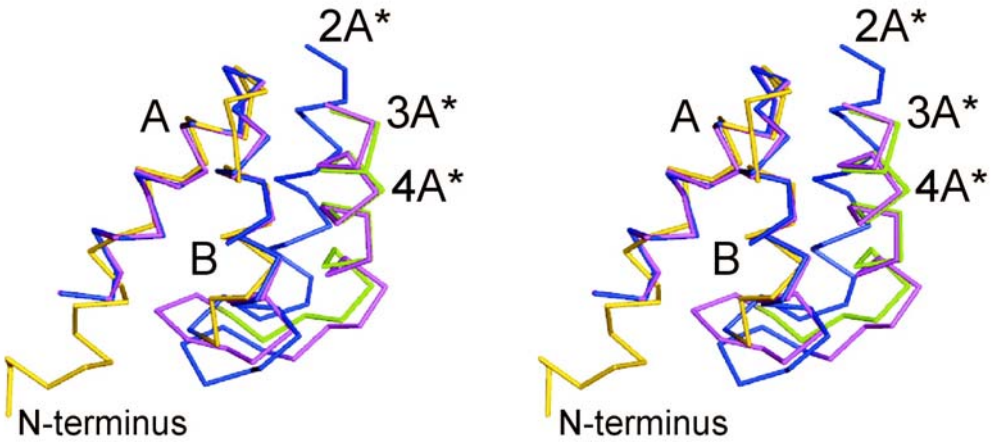


FIG. 6. **Superposition of helices A, B and A* of HcpB, where A* is helix A of the next repeat.** Note that interaction of helix 1B of the first repeat (indicated as yellow helix B) with helix 2A of the second repeat (indicated as blue helix 2A*) differs from the other inter-repeat interactions. When comparing single repeats, helices A and B are super-imposable. The superposition was calculated based on the coordinates of helices A and B of each repeat and neglecting the coordinates of helix A* of the subsequent repeat. Color code is as in Fig. 1a.

Oxidative refolding *in vitro* of HcpB proceeds without apparent reshuffling of non-native disulfide bonds. If there are "wrong" intermediates, which we cannot strictly rule out, these are very minor species escaping detection under our experimental conditions. Absence of incorrect disulfide bonds indicates that some partial repeat structure forms already before cysteine oxidation begins. Since, as outlined above, repeat 1 is essentially unfolded before cysteine-oxidation, early structuring of repeats 2, 3 and 4 seems to prevent wrong cysteine pairings by keeping non-consecutive cysteines separated. Helix structure appears immediately after rapidly removing GdmCl and DTT from reduced and unfolded HcpB. Fully reduced HcpB in the absence of chemical denaturant has approximately 50% of the helix content of the native protein (Fig. 4), in agreement with a 41% helix content calculated for the sequence of HcpB (25). Thus, one may envisage formation of the stable folding intermediate $HcpB_{3SS}$ as an inherent consequence of early structure formation within three of the four repeats. That the N-terminal repeat 1 does not fold before cysteine oxidation may be due to its peculiar structure vis-à-vis the remaining part of the molecule.

Conclusion – We found a rather simple *in vitro* mechanism of disulfide formation in a repeat protein featuring one disulfide bond in each repeat. Disulfide formation in the N-terminal repeat motif, which contributes most to the overall stability, is the rate-limiting step and proceeds from a stable folding intermediate in which three disulfides have been correctly formed. There is no obvious formation of non-native disulfides. Oxidative folding of HcpB in the periplasm of *E. coli* substantiates the compensatory roles played by the DsbA/DsbB and DsbC/DsbD systems in the folding of disulfide containing proteins in prokaryotes.

REFERENCES

1. Andrade, M. A., Perez-Iratxeta, C., and Ponting, C. P. (2001) *J. Struct. Biol.* **134**, 117-131
2. Amstutz, P., Binz, H. K., Parizek, P., Stumpp, M. T., Kohl, A., Grutter, M. G., Forrer, P., and Pluckthun, A. (2005) *J. Biol. Chem.* **280**, 24715-24722
3. Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P., Grutter, M. G., and Pluckthun, A. (2004) *J. Mol. Biol.* **22**, 575-582
4. Groves, M. R., and Barford, D. (1999) *Curr. Opin. Struct. Biol.* **9**, 383-389
5. Cao, P., McClain, M. S., Forsyth, M. H., and Cover, T. L. (1998) *Infect. Immun.* **66**, 2984-2986
6. Mittl, P. R. E., Luthy, L., Hunziker, P., and Grutter, M. G. (2000) *J. Biol. Chem.* **275**, 17693-17699
7. Graham, D. Y., and Yamaoka, Y. (1998) *Helicobacter* **3**, 145-151
8. McGee, D. J., and Mobley, H. L. (1999) *Curr. Topics Microbiol. Immunol.* **241**, 155-180
9. Krishnamurthy, P., Parlow, M. H., Schneider, J., Burroughs, S., Wickland, C., Vakil, M. B., Dunn, B. E., and Phadnis, S. H. (1999) *J. Bacteriol.* **181**, 5107-5110
10. Deml, L., Aigner, M., Decker, J., Eckhardt, A., Schutz, C., Mittl, P. R. E., Barabas, S., Denk, S., Knoll, G., Lehn, N., and Schneider-Brachert, W. (2005) *Infection & Immunity* **73**, 4732-4742
11. Luthy, L., Grutter, M. G., and Mittl, P. R. E. (2002) *J. Biol. Chem.* **277**, 10187-10193
12. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P., and Pluckthun, A. (2003) *J. Mol. Biol.* **332**, 489-503
13. Main, E. R. G., Stott, K., Jackson, S. E., and Regan, L. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5721-5726
14. Devi, S. V., Binz, K. H., Stumpp, M. T., Plückthun, A., Bosshard, H. R., and Jelesarov, I. (2004) *Protein Sci.* **13**, 2864-2870
15. Collet, J. F., and Bardwell, J. C. (2002) *Mol. Microbiol.* **44**, 1-8
16. Sevier, C. S., and Kaiser, C. A. (2002) *Nature Rev. Mol. Cell Biol.* **3**, 836-347
17. Bader, M. W., Xie, T., Yu, C. A., and Bardwell, J. C. (2000) *J. Biol. Chem.* **275**, 26082-26088
18. Berkmen, M., Boyd, D., and Beckwith, J. (2005) *J. Biol. Chem.* **280**, 11387-11394
19. Anfinsen, C. B., Haber, E., Sela, M., and White, F. H. (1961) *Proc. Natl. Acad. Sci. U. S. A.* **47**, 1309-1314
20. Ruoppolo, M., Pucci, P., and Marino, G. (2005) in *Protein Folding Handbook* (Buchner, J., and Kiefhaber, T., eds) Vol. 2, pp. 946-964, Wiley-VHC, Weinheim
21. Hennecke, J., Sebbel, P., and Glockshuber, R. (1999) *J. Mol. Biol.* **286**, 1197-1215
22. Tomb, J. F., White, O., Kerlavage, A. R., Clayton, R. A., Sutton, G. G., Fleischmann, R. D., Ketchum, K. A., Klenk, H. P., Gill, S., Dougherty, B. A., Nelson, K., Quackenbush, J., Zhou, L., Kirkness, E. F., Peterson, S., Loftus, B., Richardson, D., Dodson, R., Khalak, H. G., Glodek, A., McKenney, K., Fitzegerald, L. M., Lee, N., Adams, M. D., Hickey, E. K., Berg, D. E., Gocayne, J. D., Utterback, T. R., Peterson, J. D., Kelley, J. M., Cotton, M. D., Weidman, J. M., Fujii, C., Bowman, C., Watthey, L., Wallin, E., Hayes, W. S., Borodovsky, M., Karp, P. D., Smith, H. O., Fraser, C. M., and Venter, J. C. (1997) *Nature* **388**, 539-547
23. Wu, J., Yang, Y., and Watson, J. T. (1998) *Protein Sci.* **7**, 1017-1028
24. Pace, C. N. (1986) *Methods Enzymol.* **131**, 266-280

25. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389-3402

FOOTNOTES

Acknowledgements – We thank Rudi Glockshuber, Ulla Grauschopf, Kathrin Müntener, Mahesh Narayan, Petra Parizek and Michael T. Stumpp for valuable suggestions and help with some of the experiments, and Sergiy Chesnov for measuring mass spectra. This work was supported in part by the Swiss National Science Foundation (grants # 3100A0-100197 to I.J., # 3100-055308 to H.R.B, #310-063794 to P.R.E.M.).

[1] The abbreviations used are: CD, circular dichroism; CDAP, 1-cyano-4-dimethylamino-pyridiniumtetrafluoroborate; DTT, dithiothreitol; ESI-MS, electrospray ionization mass spectrometry; GdmCl, guanidinium chloride; GSH and GSSG, reduced and oxidized glutathione, respectively; Hcp, *Helicobacter pylori* cysteine-rich protein; HcpB, HP0336 gene product of the Hcp family; $HcpB_{1SS}$, $HcpB_{2SS}$, $HcpB_{3SS}$, folding intermediates with 1, 2 and 3 disulfide bonds, respectively; $HcpB_{\Delta 1}$, truncated HcpB lacking the first repeat motif; IPTG, isopropyl-β-D-thiogalactopyranoside; MALDI-TOF, matrix assisted laser desorption/ionization – time of flight mass spectrometry; MRE, mean ellipticity per mole of residue; Ni-NTA, nickel-nitrilotriacetic acid; RP-HPLC, reverse phase high performance liquid chromatography; TFA, trifluoroacetic acid

[2] HcpB expressed from pDsbA-HcpB in *E. coli* contains a 5-His tag, hence the mass difference

3.6 Additional experiments on the folding and stability of HcpB

In this section I describe experiments on the time course of the reductive unfolding of native HcpB (section 3.6.1) and on the thermodynamic stability of HcpB and its derivatives followed by thermal melting and denaturant unfolding (sections 3.6.2 – 3.6.5).

3.6.1 Reductive unfolding of HcpB *in vitro*

Reductive unfolding (opening of disulfide bonds) of HcpB was monitored *in vitro* at neutral pH (Scheme 3.2). The protein unfolds via the major intermediate ($HcpB_{3SS}$) which has reduced cysteines in the first repeat and oxidized, natively linked disulfide bridges in the other three repeats. Interestingly, this major intermediate is the same species that forms during oxidative refolding of HcpB (see Fig. 3 and Table I in section 3.5). The time course of reductive unfolding of HcpB monitored by RP-HPLC is shown in figure 3.6.

Native HcpB (0.5 mg/mL) in 20 mM sodium phosphate,
100 mM KCl, pH 7.0

↓

Add 15 mM DTT (final concentration)

↓

Aliquots removed at regular time intervals and reaction
quenched by one volume of 8% TFA

↓

Quenched aliquots chromatographed
by RP-HPLC and lyophilized

↓

Lyophilized fractions (0.5 mg/mL) cyanylated by 0.2 M CDAP
at acidic pH for 30 minutes

↓

Cyanylated material rechromatographed by RP-HPLC
and collected peaks split in two aliquots

↓

One aliquot subjected to mass spectrometry, the other to tryptic digestion
followed by MALDI-TOF (overnight incubation at HcpB/enzyme = 100/1; 20
mM Tris-HCl, 50 mM NaCl, pH 7.5)

Scheme 3.2. Experimental details of reductive unfolding of native HcpB.

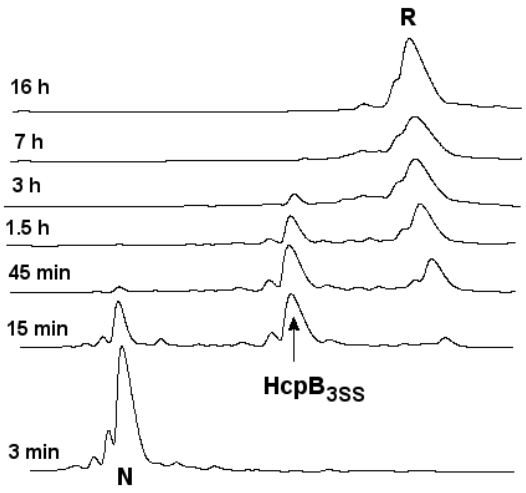


Figure 3.6. Time course of reductive unfolding of HcpB. HPLC-traces of aliqouts removed after quenching by TFA at different time intervals are shown. Peaks N (native), $HcpB_{3SS}$ and R (reduced) were collected for analysis. Minor peaks to the left and right of N have the same mass as N; the shoulder preceding R has the same mass as R, and the shoulder preceding $HcpB_{3SS}$ has the same mass as $HcpB_{3SS}$.

Masses of the cyanylated peaks N, R and $HcpB_{3SS}$ and of the peptides after tryptic digestion are exactly the same as those observed during oxidative refolding described in the manuscript in Section 3.5. It is difficult to explain why the same major intermediate was seen during reductive unfolding. One might speculate that $HcpB_{3SS}$ forms first because the first disulfide bond is most easily accessible to the reductant DTT. This is not confirmed by the crystal structure of HcpB in which the first disulfide bond is completely buried (0% solvent exposed). As both oxidative refolding and reductive unfolding of HcpB produce the same major intermediate, one may explain the major intermediate $HcpB_{3SS}$ by an opened/closed equilibrium of the first repeat. Opening would be first in unfolding and last in folding. However, we are aware that this is a strenuous argument.

3.6.2 Thermal melting monitored by CD spectroscopy

Temperature-induced unfolding curves of native HcpB recorded by circular dichroism spectroscopy are shown in Fig. 3.7. Thermal unfolding of HcpB is reversible to only ~60% at pH

>5.5. For that reason the stability of HcpB at neutral pH can not be reliably estimated from thermal melting experiments. At lower pH reversibility is higher than 85% and thermal melting provides information about ΔG_{unf} and the unfolding mechanism.

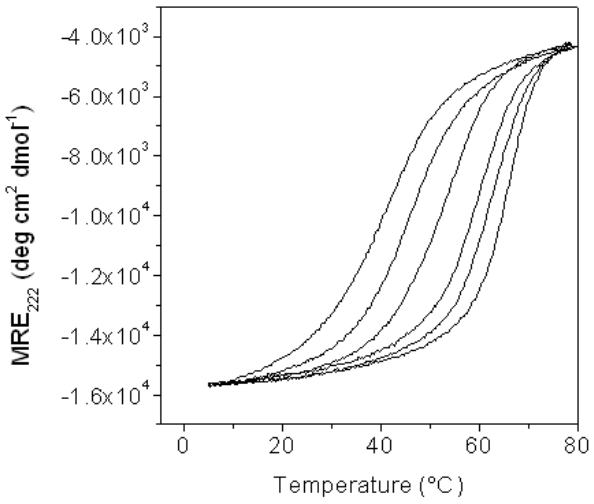


Figure 3.7. Thermal melting curves from CD spectroscopy. Change of MRE_{222} upon continuous heating at 1 deg min^{-1} at pH 2, 2.4, 3, 3.5, 4 and 5 (left to right). Protein concentration was 50 µM. Buffer: 7.5 mM each of boric, citric and phosphoric acid adjusted to the desired pH with KOH and to an ionic strength of 100 mM with KCl.

The T_m of the protein is highest at pH 5.0. Below 10 °C and above 70 °C MRE_{222} is very similar between pH 2 and pH 5, indicating that the helical content of the native state and the structure of the denatured state are not appreciably influenced by pH. However, the mid-point of thermal unfolding shifts by 25 °C between pH 2 and pH 5 (Table 3.2). Hence, acidic pH significantly destabilizes the protein.

3.6.3 Isothermal unfolding with urea

The stability of HcpB at pH 5.0 was assessed by urea unfolding at different temperatures following the change of the CD minimum at 222 nm (Fig. 3.8). All transitions were fully reversible and the protein remained fully oxidized throughout unfolding. Interestingly, we observed a significant decrease in ellipticity at high urea concentrations and increasing temperature, indicating temperature-induced structural variation of the urea-denatured state.

Another peculiarity of the unfolding curves is that MRE_{222} decreases slightly but reproducibly between 0 and 0.5 M urea. This effect is possibly caused by small structural rearrangements due to protein-urea interactions.

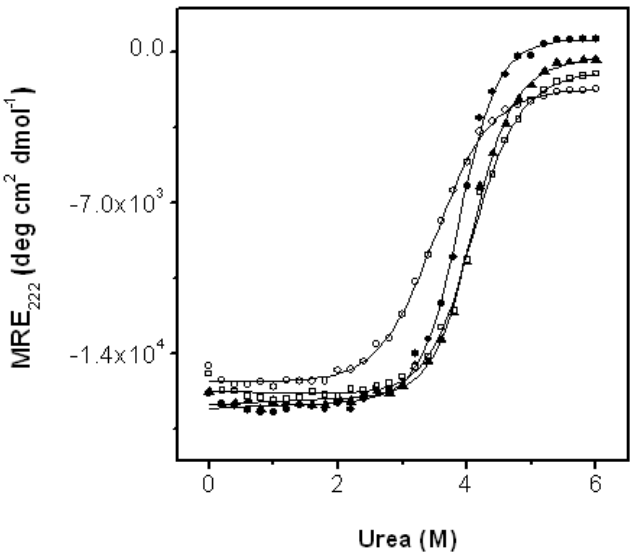


Figure 3.8. Urea-induced unfolding of HcpB. Mean ellipticity per residue at 222 nm (MRE_{222}) of HcpB (20 μM) was measured as a function of increasing urea concentrations at pH 5 and 5 °C (closed circles), 15 °C (triangles), 25 °C (squares) and 40 °C (open circles). Solid lines are best non-linear fits.

The non-linearity at low urea concentration was neglected and the data were modeled by a two-state transition between native and unfolded protein. The linear extrapolation procedure was used to calculate the Gibbs free energy of unfolding, ΔG_{unf}, at zero urea (Table 3.1). The $[urea]_{0.5}$ and m-values were very similar up to 25 °C and decreased at higher temperatures. The variation of ΔG_{unf} (~29 kJ mol^{-1}) between 5 and 25 °C is lower than the experimental error. Significant destabilization occurs at 40 °C (ΔG_{unf} = 19.6 kJ mol^{-1}).

3.6.4 Thermal melting monitored by differential scanning calorimetry (DSC)

Thermal melting at pH 2, 3, 4 and 5 was also followed by DSC (Fig. 3.9). The partial specific heat capacity at 25 °C is close to the values measured for typical globular proteins but exhibits a steeper increase with temperature: 8.5×10^{-3} J g^{-1} K^{-2} at pH 5 to 13.8×10^{-3} J g^{-1} K^{-2} at pH 2, as

compared to the reference value of $(6.7 \pm 1)\times10^{-3}$ J g^{-1} K^{-2} [32]. The transitions are relatively broad, especially at low pH.

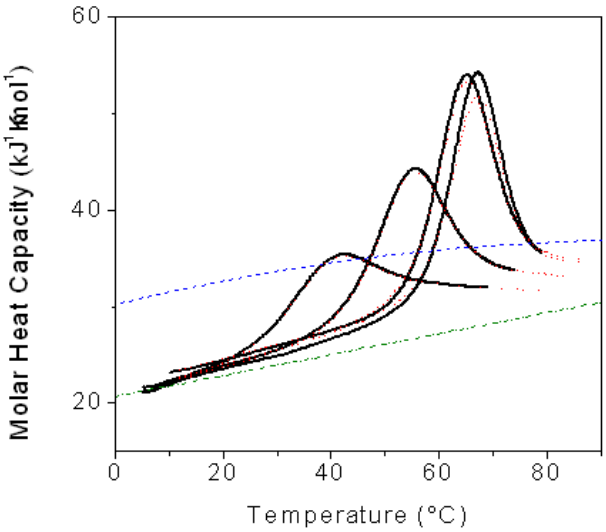


Figure 3.9. DSC traces from pH 2 to pH 5 (left to right). The change of the molar heat capacity upon continuous heating at 1 deg min^{-1} is shown. Protein concentration was 150 µM. The first heating was stopped when the protein was unfolded > 95% (continuous black lines). After cooling, the heating was repeated to higher temperatures to collect data on the heat capacity of the unfolded protein by dotted lines (red). The predicted heat capacity for complete unfolding is represented by the dashed line (blue). The heat capacity of a typical globular domain of the size of HcpB is shown by the dash-dotted line (green).

The heat capacity of the denatured state is lower than the predicted heat capacity of a fully solvated polypeptide chain with the amino acid composition of HcpB. The apparent heat capacity increment in the melting zone is ΔC_p = is 4 ± 0.3 kJ K^{-1} mol^{-1} (mean from four DSC traces). Kirchoff plots of ΔH_m vs. T_m were used to obtain ΔC_P (in units of kJ K^{-1} mol^{-1}). Values of ΔC_P were calculated from three different sets of enthalpy values: ΔC_P = 7.3 ± 1.1 from ΔH_m^{cal} (calorimetric, model-independent enthalpy), ΔC_P = 4.6 ± 0.7 from $\Delta H_m^{vH,DSC}$ (van't Hoff enthalpy from DSC), ΔC_P = 4.9 ± 1.5 from $\Delta H_m^{vH,CD}$ (enthalpy from CD melting).

It has been argued that the most reliable estimate for the unfolding enthalpy is the weighted average of ΔH_m^{cal} and $\Delta H_m^{vH,DSC}$, which is calculated as $\Delta H_{WA} = 0.35\times \Delta H_m^{cal} + 0.65 \times \Delta H_m^{vH,DSC}$ [33]. ΔC_p calculated from the plot of ΔH_{WA} vs. T_m is 5.6 ± 0.1 kJ K^{-1} mol^{-1}. All these estimates

are considerably lower than ΔC_p = 8–9 kJ K^{-1} mol^{-1} expected for a 16 kDa protein obeying the "ideal" unfolding behavior [34].

Table 3.1: ΔG_{unf} and m values from urea unfolding curves at pH 5 obtained at different temperatures. Errors of ΔG_{unf} and m are estimated as 3 kJ mol^{-1} and 0.8 kJ mol^{-1} M^{-1}, respectively.

T (°C)	$[Urea]_{0.5}$ (M)	ΔG_{unf} (kJ mol^{-1})	m (kJ mol^{-1} M^{-1})
5	3.8	28.7	7.8
15	3.8	28.9	7.2
25	3.7	28.8	7.0
40	3.4	19.6	6.4

We conclude that at the on-set of the main unfolding event, HcpB is significantly hydrated and/or quite flexible in comparison to the low-temperature folded state. This follows also from the per-residue unfolding enthalpy (2.1 kJ mol res^{-1}) and entropy (6.3 J K^{-1} mol res^{-1}) calculated at 65 °C, which both are among the lowest values that have been measured for proteins [34]. According to the DSC data, the main unfolding transition of HcpB is rather cooperative between pH 3 and 5, as judged by the ratio of the model-independent calorimetric to the model-dependent van't Hoff enthalpy, $\Delta H_m^{vH,DSC} / \Delta H_m^{cal}$, which is close to unity (Table 3.2). At pH 2, $\Delta H_m^{vH,DSC} / \Delta H_m^{cal}$ is close to two. Since melting is perfectly reversible at pH 2 and the protein is monomeric, the reason for the ratio of 2 is not clear. However, the fraction of unfolded protein, f_U, changes in a non-parallel fashion with temperature when calculated from CD or DSC data (Fig. 3.10).

Table 3.2: Thermodynamic parameters describing the thermal melting of HcpB by CD and DSC studies at different pH.

pH	T_m (°C) CD	T_m (°C) DSC	ΔH_m^{vH} (kJ mol^{-1}) DSC	ΔH_m^{cal} (kJ mol^{-1}) DSC	$\Delta H_m^{vH}/\Delta H_m^{cal}$
2	39.3	40.4	182	94	1.93
2.4	45.2				
3	51.7	53.3	216	225	0.96
3.5	58.7				
4	61.6	64.4	283	287	0.98
5	64.8	66.3	304	280	1.08

The temperature of 50% unfolding ($f_U = 0.5$) is systematically higher in the DSC experiment. It is likely, therefore, that melting of the secondary structure precedes the disruption of gross packing interactions. A further indication that unfolding of HcpB deviates from the two-state model comes from a global analysis of the data according to the Gibbs-Helmholtz equation. The unfolding free energy measured at different temperatures can be combined with $\Delta G_{unf} = 0$ at the mid-point of the thermal unfolding transition, T_m, to calculate the unfolding enthalpy at T_m and ΔC_p. At pH 5, a best fit was obtained with $\Delta C_p = 7 \pm 0.4$ kJ K^{-1} mol^{-1}. However, the calculated enthalpy was 400 kJ mol^{-1} (pH 5), in significant disagreement with the enthalpy obtained from the analysis of individual thermal melting experiments. Urea-induced unfolding predicts a higher stability than thermal melting.

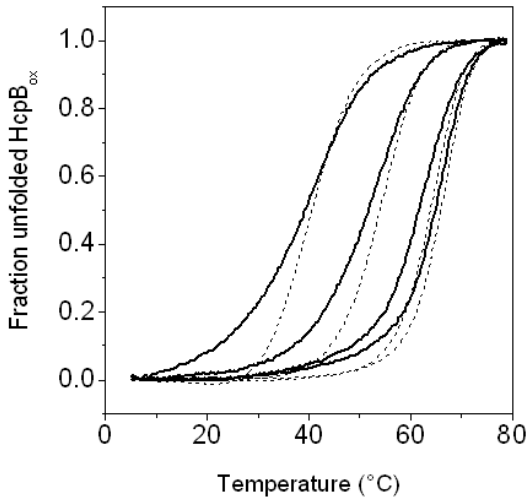


Figure 3.10. Fraction of unfolded HcpB as a function of temperature at pH 2, 3, 4, and 5 (continuous lines, left to right, calculation based on the data of Fig 3.7). Dashed lines represent fraction unfolded HcpB calculated from the DSC data of Fig. 3.9.

In conclusion, unfolding of native HcpB deviates from a two-state model and involves intermediates. The temperature-induced denatured state of HcpB is only partially hydrated, which could be due to disulfide bridges keeping the repeating motifs partially intact. During unfolding the helices melt first followed by the global unfolding event. The same we have observed for an ankyrin repeat protein [35].

3.6.5 Contribution of disulfide bridges to stability

Further to the urea unfolding experiments reported in Fig. 5 of section 3.5, we also performed thermal CD melting experiments at pH 7 with native HcpB, disulfide reduced HcpB ($HcpB_{red}$), the major folding intermediate $HcpB_{3SS}$ and the truncated protein $HcpB_{\Delta 1}$ (Fig. 3.11) to reveal the stabilizing role of disulfide bonds.

The apparent melting temperature of fully reduced HcpB is 39 °C, that of folded HcpB is 61 °C. Midpoint temperatures for the major intermediate $HcpB_{3SS}$ and the truncated protein $HcpB_{\Delta 1}$ are similar, about 56 °C. From this follows that the disulfide bridges significantly stabilize HcpB.

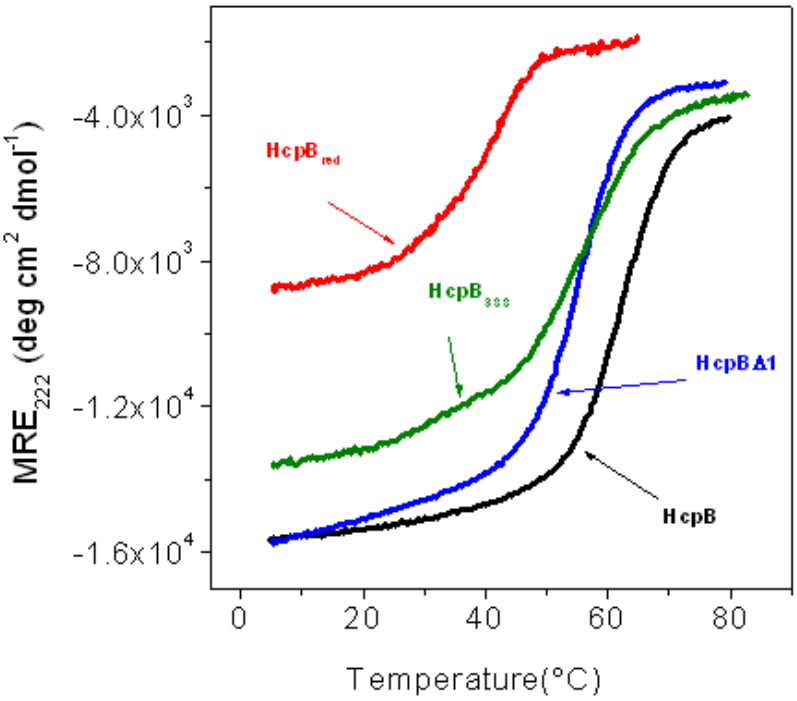


Figure 3.11. Thermal melting followed by CD at pH 7. MRE_{222} recorded with 50 µM protein at a heating rate of 1 deg min^{-1}. Native HcpB (black), $HcpB_{\Delta 1}$ (blue), $HcpB_{3SS}$ (green), reduced HcpB (red). Buffer: 20 mM sodium phosphate, 100 mM KCl, pH 7.0.

3.7 References

1. Warren, J.R. and B. Marshal, (1983). Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet* **1,** 1273-5.
2. Forman, D., (1991). Helicobacter pylori infection: a novel risk factor in the etiology of gastric cancer. *J Natl Cancer Inst.* **83,** 1702-3.
3. Blaser, M.J., (1990). Helicobacter pylori and the pathogenesis of gastroduodenal inflammation. *J Infect Dis.* **161,** 626-33.
4. Blaser, M.J., (1995). The role of Helicobacter pylori in gastritis and its progression to peptic ulcer disease. *Aliment Pharmacol Ther.* **9 Suppl 1,** 27-30.
5. Parsonnet, J., et al., (1991). Helicobacter pylori infection in intestinal- and diffuse-type gastric adenocarcinomas. *J Natl Cancer Inst.* **83,** 640-3.
6. Parsonnet, J., et al., (1991). Helicobacter pylori infection and the risk of gastric carcinoma. *N Engl J Med.* **325,** 1127-31.
7. Alm, R.A., et al., (1999). Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen Helicobacter pylori. *Nature* **397,** 176-80.
8. Tomb, J.F., et al., (1997). The complete genome sequence of the gastric pathogen Helicobacter pylori. *Nature* **388,** 539-47.

9. Letunic, I., et al., (2002). Recent improvements to the SMART domain-based sequence annotation resource. *Nucleic Acids Res.* **30,** 242-244.
10. Decker, J., et al., (2001). Identification and characterisation of a 24 kDa Helicobacter pylori extracellular protein that induces high levels of IFN-gamma in mouse splenocytes.*Int. J. Med. Microbiol.* **291,** 27-27.
11. Krishnamurthy, P., et al., (1999). Identification of a novel penicillin-binding protein from Helicobacter pylori. *J Bacteriol.* **181,** 5107-10.
12. Luthy, L., M.G. Grutter, and P.R. Mittl, (2002). The crystal structure of Helicobacter pylori cysteine-rich protein B reveals a novel fold for a penicillin-binding protein. *J Biol Chem.* **277,** 10187-93.
13. Mittl, P.R., et al., (2000). The cysteine-rich protein A from Helicobacter pylori is a beta-lactamase. *J Biol Chem.* **275,** 17693-9.
14. Luthy, L., M.G. Grutter, and P.R. Mittl, (2004). The crystal structure of Helicobacter cysteine-rich protein C at 2.0 A resolution: similar peptide-binding sites in TPR and SEL1-like repeat proteins. *J Mol Biol.* **340,** 829-41.
15. Andrade, M.A., C. Perez-Iratxeta, and C.P. Ponting, (2001). Protein repeats: structures, functions, and evolution. *J Struct Biol.* **134,** 117-31.
16. Kobe, B. and A.V. Kajava, (2000). When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem Sci.* **25,** 509-15.
17. Das, A.K., P.W. Cohen, and D. Barford, (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *Embo J.* **17,** 1192-9.
18. Creighton, T.E., (1991). Characterizing intermediates in protein folding. *Curr Biol.* **1,** 8-10.
19. Wedemeyer, W.J., et al., (2000). Disulfide bonds and protein folding. *Biochemistry* **39,** 4207-16.
20. Creighton, T.E., (1986). Disulfide bonds as probes of protein folding pathways. *Methods Enzymol.* **131,** 83-106.
21. Creighton, T.E., (1984). Disulfide bond formation in proteins. *Methods Enzymol.* **107,** 305-29.
22. Huggins, C., D.F. Tapley, and E.V. Jensen, (1951). Sulphydryl-disulphide relationships in the induction of gels in proteins by urea. *Nature* **167,** 592-3.
23. Weissman, J.S. and P.S. Kim, (1991). Reexamination of the folding of BPTI: predominance of native intermediates. *Science* **253,** 1386-93.
24. Rothwarf, D.M. and H.A. Scheraga, (1993). Regeneration of bovine pancreatic ribonuclease A. 1. Steady-state distribution. *Biochemistry* **32,** 2671-9.
25. Goto, Y. and K. Hamaguchi, (1981). Formation of the intrachain disulfide bond in the constant fragment of the immunoglobulin light chain. *J Mol Biol.* **146,** 321-40.
26. Torella, C., et al., (1994). Analysis of RNase A refolding intermediates by electrospray/mass spectrometry. *FEBS Lett.* **352,** 301-6.
27. Chang, J.Y., (1996). The disulfide folding pathway of tick anticoagulant peptide (TAP), a Kunitz-type inhibitor structurally homologous to BPTI. *Biochemistry* **35,** 11702-9.
28. Wu, J. and J.T. Watson, (1997). A novel methodology for assignment of disulfide bond pairings in proteins. *Protein Sci.* **6,** 391-8.
29. Chang, J.Y., (1993). Identification of productive folding intermediates which account for the flow of protein folding pathway. *J Biol Chem.* **268,** 4043-9.
30. Bardwell, J.C., K. McGovern, and J. Beckwith, (1991). Identification of a protein required for disulfide bond formation in vivo. *Cell* **67,** 581-9.
31. Hiniker, A. and J.C. Bardwell, (2003). Disulfide bond isomerization in prokaryotes. *Biochemistry* **42,** 1179-85.
32. Gomez, J., et al., (1995). The heat capacity of proteins. *Proteins* **22,** 404-12.
33. Haynie, D.T. and E. Freire, (1994). Estimation of the folding/unfolding energetics of marginally stable proteins using differential scanning calorimetry. *Anal Biochem.* **216,** 33-41.

34. Robertson, A.D. and K.P. Murphy, (1997). Protein Structure and the Energetics of Protein Stability. *Chem Rev.* **97,** 1251-1268.
35. Devi, V.S., et al., (2004). Folding of a designed simple ankyrin repeat protein. *Protein Sci.* **13,** 2864-70.

4 Different Effects of Urea and Guanidinium Chloride on the Free Energy of Unfolding of a Protein

(An attempt at solving a longstanding problem)

4.1 Introduction

The linear extrapolation model (LEM) – The free energy of unfolding of a protein is the difference between the free energies of the folded and unfolded states: $\Delta G_U = G_U - G_F$, where G_U and G_F are the absolute free energies of the unfolded and folded state, respectively, and ΔG_U is the free energy that is necessary to unfold the protein. In other words, ΔG_U is a measure of the protein's thermodynamic stability. While the absolute free energies, G_U and G_F, are not open to experiment, there are several ways to determine ΔG_U. A popular one is the linear extrapolation method, LEM for short (1). It provides an estimate of ΔG_U in a very simple, albeit indirect way: ΔG_U is measured in the presence of increasing concentrations of a chemical denaturant, [den], and ΔG_U^W, the free energy of unfolding in plain buffer, is obtained by linear extrapolation to zero denaturant concentration according to

$$\Delta G_U = \Delta G_U^W - m\,[den] \qquad (4.1)$$

The slope m can be regarded as the "strength of the denaturant effect" on the native conformation. Larger the m, the more sensitive the protein to chemical denaturation. Urea and GdmCl are the most commonly used denaturants.

4.1.1 The problem

To be generally applicable, linear extrapolation should provide the same value of ΔG_U^W independent of the nature of the denaturant. However, for the same protein different ΔG_U^W-values are sometimes extrapolated from urea and GdmCl unfolding data, respectively. In LEM, the properties of the folded and denatured state in the transition zone of unfolding, that is, in the presence of sometimes high concentrations of denaturant, are projected to zero denaturant concentration. If urea and GdmCl alter the thermodynamic character of the folded and denatured states in different ways, extrapolation to zero denaturant concentration may not yield the same free energy of unfolding. GdmCl, but not urea, acts as a dissociated salt and screens electrostatic

interactions. The screening effect of GdmCl is perhaps the most important reason why chemical unfolding may occasionally yield inconsistent ΔG_U^W-values. We know of only a few reports in which the effect of urea and GdmCl on both the kinetics and equilibrium thermodynamics of protein folding have been compared (2-6).

We have chosen the dimeric, disulfide-linked leucine zipper AB_{SS} to test how the nature of the denaturant influences the magnitude of ΔG_U^W. In particular, we compare the effect of urea and GdmCl on the equilibrium value of ΔG_U^W calculated according to eq 4.1 and on the value of ΔG_U^W deduced from the rates of folding and unfolding according to

$$\Delta G_U^W = -RT\ln(k_f/k_u) \qquad (4.2)$$

Here, k_f and k_u are the first order rate constants of folding and unfolding, respectively. They again are obtained by linear extrapolation from values of k_f and k_u measured in the presence of different concentrations of denaturant.

4.1.2 Previous work on electrostatic effects on the stability of leucine zippers

Electrostatic interactions between oppositely charged side chains, called "salt bridges", are typical features of proteins. Salt bridges are commonly assumed to stabilize the native protein. However, experimental and computational analyses indicate that salt bridges are not necessarily stabilizing (7). Leucine zippers (or coiled coils in general) are good models to study electrostatic interactions in proteins because they feature salt bridges between adjacent helices, and these can easily be manipulated by mutation (8-10). In the simplest case, leucine zippers consist of two α-helical peptides wound around each other and held together by hydrophobic packing along the dimer interface. The amphipathic coiled coil structure originates from a 7-residue sequence motif, $(abcdefg)_n$, repeating every two α-helical turns. Hydrophobic residues dominate at positions *a* and *d* of each heptad while positions *e* and *g* are often charged.

Marti and Bosshard used designed leucine zippers to probe the effect of electrostatic interactions to protein stability (11-14). The work described in this Chapter is based on leucine zipper AB_{SS} (13). Figure 4.1 shows a helical wheel representation of AB_{SS}, which is composed of an acidic A-chain and a basic B-chain. The A- and B-chains by themselves are unfolded because of charge-charge repulsions and because the hydrophobic residues in positions *a* and *d* of a single α-helix

are exposed to the aqueous environment. When combined, A- and B-chains associate spontaneously to a heterodimeric coiled coil, a phenomenon dubbed “peptide Velcro” (8).

Ac-*E*-**V**AQ**L***E*K*E*-**V**AQ**A***E*A*E*-**N**YQ**L***E*Q*E*-**V**AQ**L***E*H*E*-**C**G-NH_2

Ac-*E*-**V**AQ**L***K*K*R*-**V**QA**L***K*A*R*-**N**YA**L***K*Q*K*-**V**QA**L***R*H*K*-**C**G-NH_2

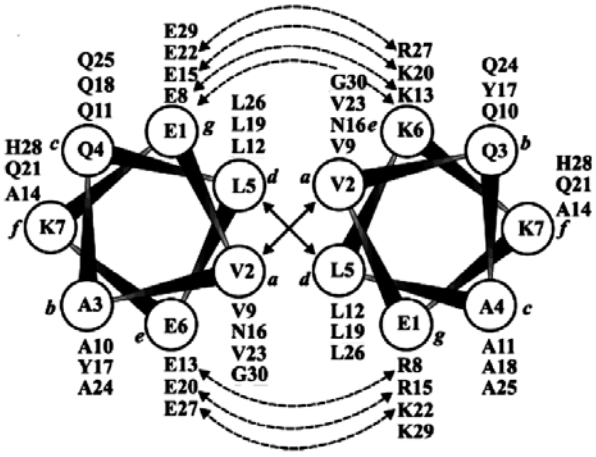


Figure 4.1 Top: Sequence of the dimeric, disulfide-linked leucine zipper AB_{SS}. Heptad repeats are separated; *a* and *d* heptad position residues are in bold, *g* and *e* heptad position residues are in italics. The C-terminal disulfide bridge is indicated by a vertical line. **Bottom: Helix wheel representation of leucine zipper AB_{SS}.** Heptad positions are denoted with lower case letters in italics. Solid arrows denote hydrophobic interaction contacts of *a* and *d* heptad position residues at the chain interface. Dashed arrows denote inter-chain salt bridge formation between *g* and *e* heptad position residues of chains A (left) and B (right). Figure adapted from (14).

In the design of AB_{SS}, the two chains have been linked by a C-terminal disulfide bridge. This facilitates both the thermodynamic and the kinetic analysis of the folding/unfolding reaction since reactions become concentration-independent if the two chains are disulfide-linked. The 3D structure of AB_{SS}, solved by ^{1}H-NMR spectroscopy (13), is shown in Figure 4.2.

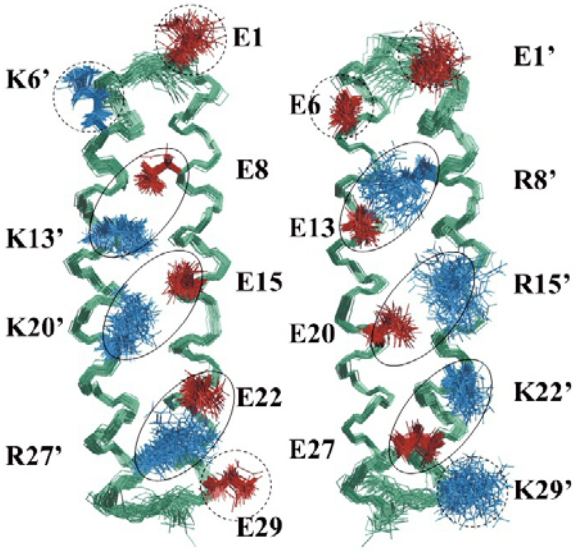


Figure 4.2 Backbone traces and selected side chain bonds of 48 AB_{SS} NMR structures superimposed by fitting backbone atoms to the mean. Acidic residues (Glu) are displayed in red and basic residues (Lys, Arg) in blue; backbone bonds are colored green; groups of salt bridges are encircled by solid lines and unpaired charges by dashed lines. "Front" and "back" views of the molecule are shown. Figure adapted from (13).

The energetic contribution of ionizable Glu and His side chains was evaluated from the residues' pK_a values in the folded and unfolded leucine zipper, respectively. The contribution of a single charge to the free energy of unfolding of the protein is proportional to the pK_a change between the folded and unfolded structure (7). Some charges were found to stabilize the folded leucine zipper, others were destabilizing.

4.1.3 Previous work on the effect of urea and GdmCl on leucine zipper AB_{SS}

The thermodynamic stability of AB_{SS} has been studied over the pH range 2–7 by chemical denaturation at constant temperature as well as by thermal unfolding (12). Urea unfolding and thermal denaturation yield similar values of ΔG_U^W. Unfolding by GdmCl, however, yields significantly different unfolding parameters. The results are summarized in Figure 4.3 taken from (12).

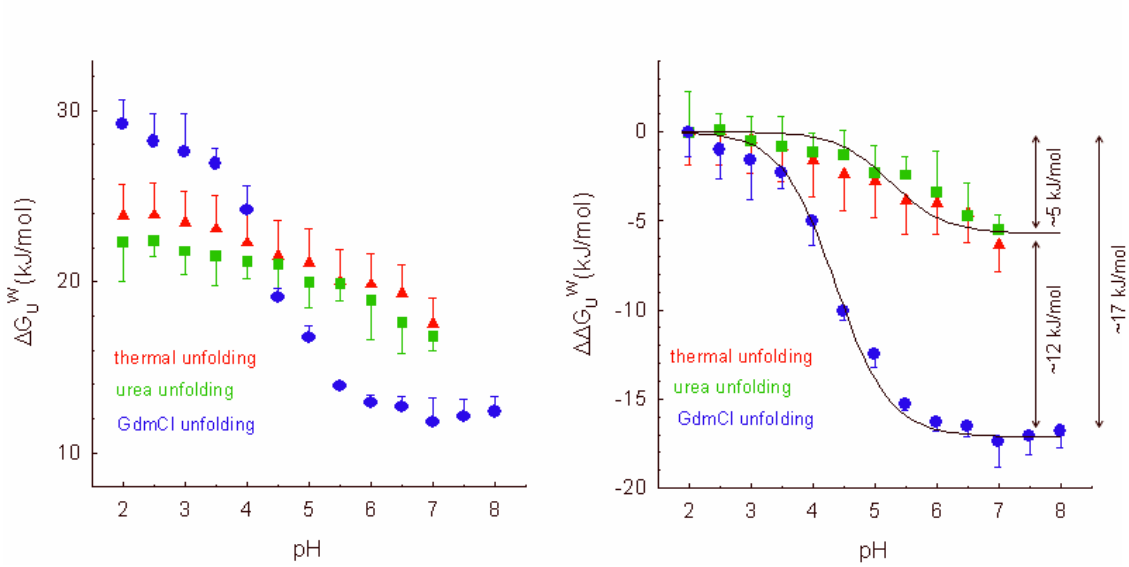


Figure 4.3 pH-stability profile of leucine zipper AB_{SS} presented as the change in the free energy of unfolding vs. pH. Left: Absolute free energy of unfolding, ΔG_U^W vs. pH. Right: Relative free energy of unfolding $\Delta\Delta G_U^W$ defined as $\Delta G_U^W(pH) - \Delta G_U^W(pH2)$ vs. pH. Figure adapted from (12).

Three features of Figure 4.3 spring to the eye:

(i) Leucine zipper AB_{SS} is more stable at low pH than at neutral pH, indicating a net unfavorable charge effect at neutrality. (The larger ΔG_U^W, the more stable the native protein.)

(ii) In absolute terms, AB_{SS} is more stable at pH 2 when ΔG_U^W is extrapolated from GdmCl unfolding. The opposite is seen at pH 7: AB_{SS} is more stable when ΔG_U^W is extrapolated from urea unfolding (left panel).

(iii) The stability difference between pH 2 and pH 7 is larger for GdmCl unfolding than for urea and thermal unfolding, ~17 vs. ~5 kJ/mol (right panel).

Observations (i) – (iii) are clear manifestations of denaturant effects. The experiments to be described in this Chapter attempt at explaining these denaturant effects by following the kinetics of folding of AB_{SS} at acidic and neutral pH.

4.1.4 Relationship between free energies of unfolding and kinetic rate constants

If GdmCl and urea alter the structures of native and unfolded AB_{SS} in different ways, this will lead to denaturant effects on the folding and unfolding rates of AB_{SS}. For example, charge screening by GdmCl might reduce unfavorable charge-charge interactions and thereby stabilize the folded or the unfolded state of AB_{SS}. Similarly, charge screening might reduce favorable charge-charge attraction and thereby reduce protein stability. As a consequence, one expects charge screening to also affect the rates of folding and unfolding. Since charge-charge interactions are pH-dependent, the kinetic effects of charge screening will be pH-dependent as well. From measuring the rates of folding, k_f, and unfolding, k_u, in the presence of increasing amounts of denaturant, the rate constants in the absence of denaturant, k_f^W and k_u^W, are linearly extrapolated according to:

$$ln k_f = ln k_f^W - m_f[den]/RT \qquad (4.3)$$

$$ln k_u = ln k_u^W + m_u[den]/RT \qquad (4.4)$$

The rate constant accessible to experiment is k_{obs}. Since $k_{obs} = k_f + k_u$, it follows:

$$ln k_{obs} = k_f^W exp(-m_f[den]) + k_u^W exp(m_u[den]) \qquad (4.5)$$

A plot according to eq 4.5 of $ln k_{obs}$ vs. the denaturant concentration is called a Chevron plot and yields values for k_f, k_u, m_f and m_u. The slopes m_u and m_f describe the dependence of the rate constant on denaturant concentration. The magnitude of m_u and m_f is an empirical measure of the change in the solvent exposed surface when the protein passes through the transition state separating the folded from the unfolded state. Since m of the equilibrium chemical unfolding experiment (eq 4.1) is proportional to the overall surface change accompanying the unfolding reaction, there follows the relationship:

$$m = RT(m_f - m_u) \qquad (4.6)$$

To be consistent, equilibrium and kinetic unfolding experiments should yield similar values of ΔG_U^W (eqs 4.1 and 4.2) and m (eqs 4.1 and 4.6).

4.2 Materials and methods

Materials – Leucine zipper AB_{SS} was chemically synthesized and purified as described (11). GdmCl and urea were of the highest purity available from Fluka. The concentration of AB_{SS} was determined by UV absorption in 6 M GdmCl using $\varepsilon_{275.3} = 1450\ M^{-1}cm^{-1}$.

Buffers – All experiments were conducted in buffer composed of phosphoric, citric and boric acid, 7.5 mM each, adjusted to the desired pH with KOH or HCl and to an ionic strength of 0.1 M with KCl. The pH of buffers containing urea or GdmCl was adjusted after adding the denaturant. The concentration of denaturant was measured by refractometry.

Stopped-Flow Measurements – Experiments were performed with the help of a π*-180 rapid mixing instrument (Applied Photophysics). The instrument had an optical path length of 10 mm and a dead time of 1–2 ms. The wavelength of detection was 225 nm and the slits were set to 4 mm. Ten to 15 firings were averaged for each kinetic trace. Raw data were analyzed with the software provided by the manufacturer.

Experiments with GdmCl – Unfolding of AB_{SS} was followed after 1:10 dilution of AB_{SS} (100 μM) with GdmCl. For measurements at pH 2, the final concentration of GdmCl was ≥5 M, and at pH 7 ≥3 M. Refolding was measured by rapidly mixing of one volume of AB_{SS} (100 μM) in 6 M GdmCl (pH 2), or in 4 M GdmCl (pH 7.0), with 10 volumes of buffer containing an appropriate concentration of denaturant.

Experiments with urea – These were performed in the same way as the experiments with GdmCl except that for unfolding, the final concentration of urea was ≥7 M at pH 2 and ≥6 M at pH 7. For refolding, AB_{SS} was pre-equilibrated with 8 M urea at pH 2 and 8 and rapidly diluted 1:10 with appropriate concentrations of urea in buffer.

Error calculation – Standard errors of extrapolated rate constants and m-values were estimated by the method of error propagation using data sets truncated by one data point at a time.

4.3 Results

Refolding rate constants were measured by rapid dilution of denaturant unfolded AB_{SS} into folding conditions. Unfolding rate constants were obtained in the same way by rapidly diluting the folded protein into GdmCl or urea solutions. The reaction was monitored from the change of the CD absorption at 222 nm, a CD minimum typical of the folded α-helix structure. The signal change could be described by a single exponential rise or decay, in agreement with a concentration-independent, mono-molecular reaction. The relaxation time, τ, of the single exponential function yields $k_{obs} = 1/\tau$. Chevron plots of the kinetic experiments are shown in Figure 4.4. Kinetic parameters extrapolated to zero denaturant concentration with the help of eq 4.5 are summarized in Table 4.1.

There is one unique and highly significant difference between the urea and GdmCl data. At pH 2, k_u^W from the GdmCl experiment is three orders of magnitude smaller than k_u^W from the analogous urea experiment. All other rate constants from the GdmCl and urea experiments are similar, or even identical within the limits of experimental uncertainty.

The main difficulty with these experiments was the very high rate of refolding and, in the case of urea, the weak dependence of unfolding on urea concentration. When AB_{SS} was diluted into low concentrations of denaturant, refolding was completed within the dead time of the instrument. Therefore, the extrapolation of the rate constants and m-values had to span over a large area for which no experimental data points are available (Fig. 4.4).

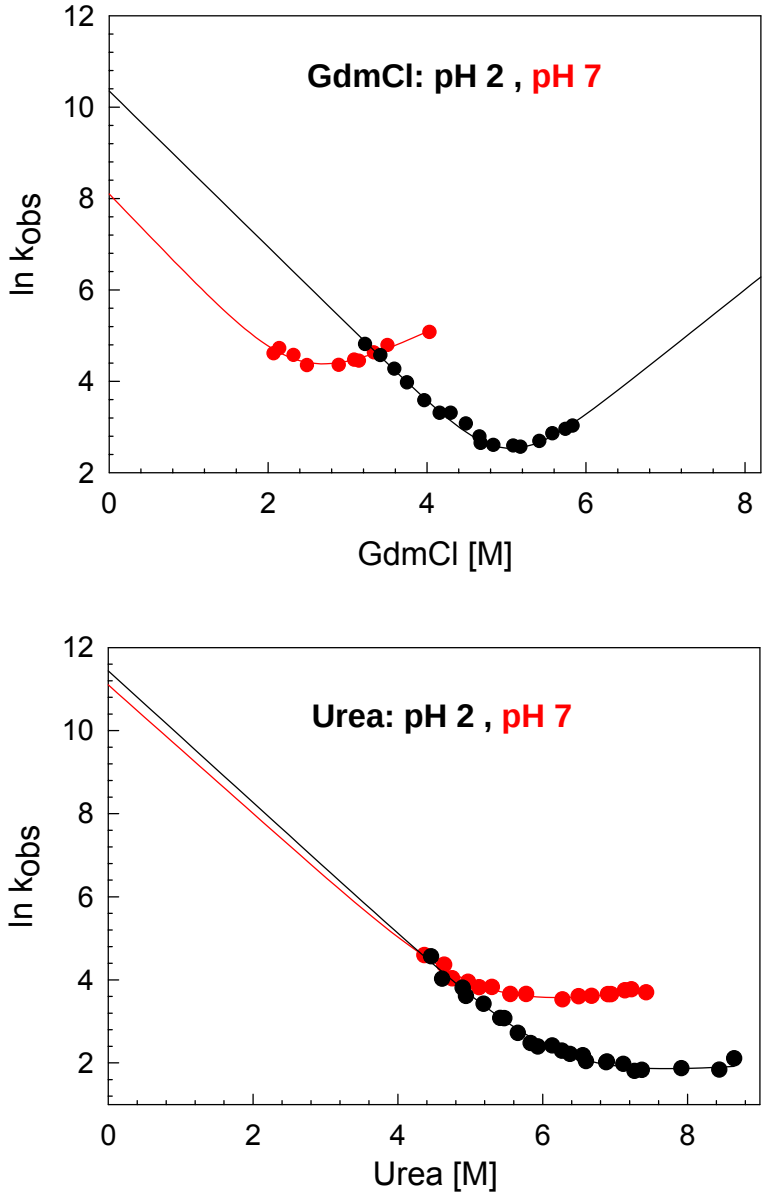


Figure 4.4 Chevron plots of folding/unfolding of AB_{SS} deduced from rapid dilution into GdmCl and out of GdmCl (top), or into urea and out of urea (bottom). Data for pH 2 and pH 7 are distinguished by color. Solid lines are best non-linear fits according to eq 4.5.

Table 4.1 Kinetic parameters extrapolated from Chevron plots of Figure 4.4 by non-linear fitting according to eq. 4.5. The standard error of the extrapolated rate constants and m-values is in the range of 20–50%. The only two significantly different rate constants are in bold.

	GdmCl				
	k_f^W (s^{-1})	k_u^W (s^{-1})	m_f (M^{-1})	m_u (M^{-1})	$\Delta G_U^{W\,a)}$ (kJ mol^{-1})
pH 2	3.1×10^4	**6×10^{-3}**	1.7	–1.4	35.8
pH 7	2.1×10^4	7.6	2.7	–0.76	18.2
	Urea				
	k_f^W (s^{-1})	k_u^W (s^{-1})	m_f (M^{-1})	m_u (M^{-1})	ΔG_U^W (kJ mol^{-1})
pH 2	3.0×10^4	**0.5**	1.3	–0.30	25.5
pH 7	2.6×10^4	6	1.3	–0.25	19.4

a) $\Delta G_U^W = -RT \ln (k_f^W/k_u^W)$

4.4 Discussion

4.4.1 Limits of the kinetic analysis

The main difficulty with the kinetic experiments described in this Chapter is in the extremely rapid rate of folding of AB_{SS}. Even at moderate denaturant concentrations, the apparent folding rate, k_{obs}, was of the order of the reciprocal dead time of the instrument, about 500–1000 s^{-1}. Therefore, refolding could not be measured below 4 M urea or below 2–3 M GdmCl (see the distribution of the experimental data points in Figure 4.4). Also, unfolding needed high denaturant concentrations. Hence, both the folding and unfolding limbs of the experimental Chevron plots are rather short. This is a serious weakness of all the kinetic experiments with AB_{SS}. It is for this reason that the project was abandoned. Nevertheless, some tentative conclusions can be drawn and a plausible explanation for the different effects of GdmCl and urea on AB_{SS} unfolding can be given.

4.4.2 Effect of denaturant on rates of folding and unfolding

The following discussion assumes the validity of eqs 4.3 – 4.5, that means, a linear relationship between the rates of refolding/unfolding and the denaturant concentration. Thus, any deviation from linearity at very low (or very high) denaturant concentration, known as a "roll-over", is excluded. With this limitation in mind, the most conspicuous result is the thousand fold lower rate of unfolding at pH 2 deduced from the GdmCl data: $k_u^W = 0.006\ s^{-1}$ (GdmCl) vs. $7.6\ s^{-1}$ (urea). This difference is the main reason why at pH 2, AB_{SS} appears twice as stable in the GdmCl experiment than in the urea experiment. Actually, the very different k_u^W-values at pH 2 constitute the only significant difference between the GdmCl-deduced and urea-deduced kinetic rate constants. This is a remarkable result. It may be interpreted by a significant stabilization of folded AB_{SS} at pH 2 through charge screening by GdmCl because a more stably folded AB_{SS} structure is expected to unfold more slowly. On the other hand, virtually unchanged rate constants at pH 7 extrapolated from the GdmCl and the urea experiment, respectively, indicate no significant charge screening by GdmCl at neutral pH. Also the rates of folding at pH 2 and pH 7 in the absence of denaturant are very similar and apparently independent of the denaturant: $k_f^W \sim (2–3)\times10^4\ s^{-1}$. We believe this is in support of a common folding mechanism and of the validity of LEM. Put differently, non-equivalence of k_f^W-values would indicate differences in the folding transition induced by GdmCl and urea, respectively.

Why does charge screening manifest itself only in different values of k_u^W at pH 2? In principal, charge screening takes effect at any pH and in both the folded and unfolded AB_{SS}. Figure 4.5 shows a very simplified scheme of charge screening effects on the conformational states of AB_{SS} at pH 2 and 7. At neutral pH, folded AB_{SS} features several salt bridges as well as a few unpaired charges (Figure 4.1). At pH 2, plus charged side chains of Lys and Arg dominate while Glu side chains are protonated (unchared). Hence, charge-effects at pH 2 are repulsive and unfavorable. We assume that folded AB_{SS} is stabilized thanks to lesser charge-charge repulsion at pH 2, and that this leads to the very slow rate of dissociation of AB_{SS} under acidic conditions. We recognize there might be other explanations. For example, GdmCl may strengthen hydrophobic interactions in folded AB_{SS}.

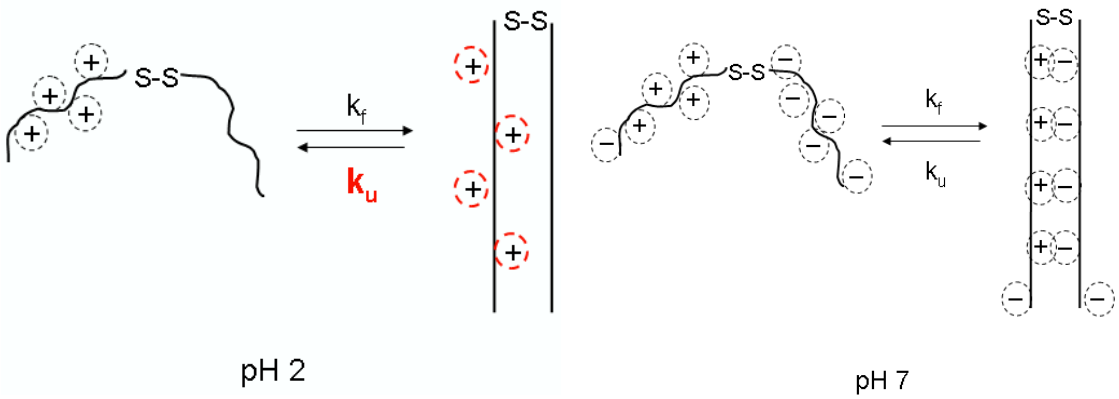


Figure 4.5 Schematic representations of charge effects in folded and unfolded AB_{SS} at pH 2 and pH 7, respectively. Charge screening is symbolized by dotted circles around charges. In this very simple scheme, only screening of full charges is considered. The only manifest kinetic effect of charge screening occurs in the folded protein at pH 2 (red dashed circles) and leads to a thousand-fold smaller value of k_u (red).

At pH 7 charge screening should weaken interactions between plus and minus charges and lower the effect of single charges (Figure 4.5). But GdmCl has no kinetic effect at pH 7. Perhaps access of the guanidinium ion to charge-charge pairs and single minus charges is sterically hindered in folded AB_{SS} at pH 7.

4.4.3 Correlation between kinetic and equilibrium parameters

Table 4.2 compares the kinetic and equilibrium thermodynamic data sets. The correlation is satisfactory. Absolute values of ΔG_U^W from equilibrium and kinetic data differ by as much as 7 kJ mol^{-1}. Unlike in the equilibrium unfolding experiment, the stability of AB_{SS} at pH 7 is very similar from both the GdmCl and urea unfolding data, 18.2 versus 19.4 kJ mol^{-1}. In equilibrium unfolding AB_{SS} is much less stable in the GdmCl experiment, 12 kJ mol^{-1} versus 17.5 kJ mol^{-1}. However, the relative changes of ΔG_U^W between pH 2 and pH 7 are remarkably similar. Thus, the change of stability between pH 2 and 7, expressed as $\Delta\Delta G_U^W$, is the same within error whether determined by equilibrium unfolding experiments or by kinetics: about 17 kJ mol^{-1} for the GdmCl experiment and about 5 kJ mol^{-1} for the urea experiment. This congruence of the equilibrium and kinetic results supports the validity of the linear extrapolation method used to analyze the very limited set of kinetic data.

Since m_u and m_f are proportional to the change in the solvent exposed surface when the protein passes through the transition state separating the folded from the unfolded state, one can define the parameter β_{TS} to indicate the position of the transition state between the folded and the unfolded state: $\beta_{TS} = -RTm_u/m$. The smaller β_{TS}, the more folded is the transition state. Values of β_{TS} calculated from the kinetic data are below 0.2, in line with a rather folded transition state. If GdmCl has any effect it is to slightly increase β_{TS}, that is, to render the transition state "less folded". However, the effect is small and probably insignificant.

Table 4.2 Comparison of data from equilibrium unfolding and kinetic experiments

	Equilibrium data (from ref. (12))			
	GdmCl		urea	
	pH 2	pH 7	pH 2	pH 7
ΔG_U^W (kJ mol^{-1})	29	12	22.5	17.5
m (kJ mol^{-1} M^{-1})	6.6	6.6	3.6	3.6
	Kinetic data (this work)			
	GdmCl		urea	
	pH 2	pH 7	pH 2	pH 7
ΔG_U^W (kJ mol^{-1})	35.8	18.2	25.5	19.4
m_f (M^{-1})	1.7	2.7	1.3	1.3
m_u (M^{-1})	–1.4	–0.76	–0.30	–0.25
$m = RT(m_f - m_u)$ (kJ mol^{-1} M^{-1})	7.7	8.6	4.0	3.8
$\beta_{TS} = -RTm_u/m$	0.18	0.09	0.08	0.07

4.4.4 Conclusion

Because of severe experimental limitations a comprehensive analysis of the folding/unfolding kinetics of AB_{SS} was not possible. Still, we can tentatively explain the previously observed higher stability of AB_{SS} extrapolated from GdmCl unfolding. The rate of unfolding of AB_{SS} at pH 2 is slowed by one thousand-fold. This is taken to indicate strong screening by GdmCl of unfavorable charge-charge repulsions at pH 2, making folded AB_{SS} significantly more stable at acidic pH.

4.5 References

1. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266-280
2. Monera, O. D., Kay, C. M., and Hodges, R. S. (1994) Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci.* **3**, 1984-1991
3. Gupta, R., Yadav, S., and Ahmad, F. (1996) Protein stability: urea-induced versus guanidine-induced unfolding of metmyoglobin. *Biochemistry* **35**, 11925-11930
4. Gupta, R., and Ahmad, F. (1999) Protein stability: Functional dependence of denaturational Gibbs energy on urea concentration. *Biochemistry* **38**, 2471-2479
5. Ibarra-Molero, B., Loladze, V. V., Makhatadze, G. I., and Sanchez-Ruiz, J. M. (1999) Thermal versus guanidine-induced unfolding of ubiquitin. An analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry* **38**, 8138-8149
6. Yang, M., Ferreon, A. C. M., and Bolen, D. W. (2000) Structural thermodynamics of a random coil protein in guanidine hydrochloride. *Proteins* **4**, 44-49
7. Bosshard, H. R., Marti, D. N., and Jelesarov, I. (2004) Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings. *J. Molec. Recogn.* **17**, 1-16
8. O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) Peptide 'Velcro': Design of a heterodimeric coiled coil. *Curr. Biol.* **3**, 658-667
9. Monera, O. D., Kay, C. M., and Hodges, R. S. (1994) Electrostatic interactions control the parallel and antiparallel orientation of alpha-helical chains in two-stranded alpha-helical coiled-coils. *Biochemistry* **33**, 3862-3871
10. Krylov, D., Barchi, J., and Vinson, C. (1998) Inter-helical interactions in the leucine zipper coiled coil dimer: pH and salt dependence of coupling energy between charged amino acids. *J. Mol. Biol.* **279**, 959-972
11. Marti, D. N., Jelesarov, I., and Bosshard, H. R. (2000) Inter-helical ion pairing in coiled coils: solution structure of a heterodimeric leucine zipper and determination of pKa values of Glu side chains. *Biochemistry* **39**, 12804-12818
12. Phelan, P., Gorfe, A. A., Jelesarov, J., Marti, D. N., Warwicker, J., and Bosshard, H. R. (2002) Salt bridges destabilize a leucine zipper designed for maximized ion pairing between helices. *Biochemistry* **41**, 2998-3008
13. Marti, D. N., and Bosshard, H. R. (2003) Electrostatic interactions in leucine zippers: thermodynamic analysis of the contributions of Glu and His residues and the effect of mutating salt bridges. *J. Mol. Biol.* **330**, 621-637
14. Marti, D. N., and Bosshard, H. R. (2004) Inverse electrostatic effect: Electrostatic repulsion in the unfolded state stabilizes a leucine zipper. *Biochemistry* **43**, 12436-12447

Publications

Devi, V. S., Binz, H. K., Stumpp, M. T., Plückthun, A., Bosshard, H. R. and Jelesarov, I. (2004) **Folding of a designed simple ankyrin repeat protein,** Protein Sci. 13: 2864-2870.

Devi, V. S., Sprecher, C. B., Hunziker, P., Mittl, P. R. E., Bosshard, H. R. and Jelesarov, I. (2006) **Disulfide formation and stability of a cysteine-rich repeat protein from *Helicobacter pylori*** accepted for publication in Biochemistry.

Poster Presentations

Folding of designed simple ankyrin repeat

a) Third International Symposium on New Trends in Structural Biology, National Center for Competence in Research, November 14–15, 2003, ETH Zurich

b) International Summer School, Karolinska Institutet, Stockholm, May 8-14, 2004

Biophysical studies of *Helicobacter pylori* cysteine-rich protein B (HcpB)

The International Network of Protein Engineering Centers, September 22–25, 2004, Brunnen, Switzerland.

Acknowledgements

The work leading to this thesis was done during my years in Prof. Hans Rudolf Bosshard's group at the Biochemistry Department, a stimulating environment to do science. The journey undertaken while pursuing my PhD was an enjoyable one and I would like to convey my heartfelt gratitude to those people who made this trip remarkable.

First of all, I want to thank **Prof. Hans Rudolf Bosshard** for mentoring and training me over the last four years in the area of protein folding. He is my supervisor who gave me a staunch support right from the beginning for my research work. He inspired me to be a life-long learner and passionate seeker of knowledge. I thank him with the deepest of respect for his support and encouragement. I enjoyed full freedom under his supervision to implement whatever ideas I had in research. Thanks also to **Prof. Markus Grutter** for co-refereeing my thesis.

To **Dr. Ilian Jelesarov**, I believe no words I could write to fully encompass the amount of gratitude I have for his supervision, moral and mental support from time to time. He afforded numerous thought provoking discussions (even on weekends), kept me to focus on important scientific questions, shaped and improved my knowledge in protein folding and tolerated my endless pestering regarding research. Thanks to **Dr. Christine Berger**. She plays different roles in my life by being my colleague, advisor, guide, friend, etc. We worked together in two projects and that was a wonderful experience which I would like to treasure throughout my lifetime. I extend my gratitude to my colleagues Dr. Daniel Marti, Dr. Susanne Cranz, Dr. Stoyan Milev, Dr. Alemaheyu Gorfe Abebe, Sasa Bjelic, Laura Andrejszki and Oksana Okhrimenko who kept the lab colorful and cheerful.

My special thanks to **Prof. Andreas Plückthun**, Dr. Kaspar Binz, Dr. Michael Tobias Stumpp, Dr. Patrick Forrer for providing me the clone of ankyrin and enabling me to work in that project. Thanks to Thomas Huber for light scattering measurements. I thank **Dr. Peer Mittl** and Dr. Lucas Luthy who provided clones for the HcpB project and kept me up-to-date about the developments in the research field of the Hcp protein family. My sincere thanks go to **Dr. Peter Hunziker** for his enthusiastic participation in the HcpB project and to Dr. Sergiy Chesnov for mass measurements. I acknowledge **Dr. Katherin Müntener** and **Prof. Antonio Baici** for teaching me western blot analysis and providing me full freedom to work in their lab. I thank all the colleagues and friends of other labs in our Department. Thanks to the management, workshop and IT administrators.

Last but not least, I am ever grateful to my **mother, father, grandmother and almighty** that are the creators of my soul and brought me up to this level. I would like to thank my Rajkumar and his family, my sister Mrs. Lalitha devi and her family, my brother Mr. Prabu and his family for their continous support and concern. Without them I dare not to start my research work and finish it successfully.

Thanks to all the above chaperones who assisted my dissertation work to fold back to its native active state.

Curriculam vitae

Name:	V. Sathya Devi
Date of Birth:	3rd November, 1977
Place of Birth:	Tiruvarur (India)
Nationality:	Indian

Education

2002-2005	Dissertation in the Department of Biochemistry, University of Zürich, Supervisor: Prof. Hans Rudolf Bosshard, Co-supervisor: PD Dr. Ilian Jelesarov. Title: "Folding of repeat proteins: Designed ankyrin repeat and cysteine-rich repeat protein from *Helicobacter pylori*".
2001-2002	Research and development officer at Hindustan Lever Research Centre (HLRC), Unilever, Bangalore, India. Project Managers: Dr. Girish Rao and Dr. Subir Das. Topic: "Delayed release of surfactants in detergent powders"
2001	Project trainee at the National Centre for Biological Sciences (NCBS) Bangalore, India. Supervisors: Dr. Sudhir Krishna and Dr. Sowdhamini. "Role of ankyrin repeat motifs in Notch mediated transformation - a structural and functional approach"
1999-2001	Master studies in Applied Chemistry at the National Institute of Technology, Trichirapalli (NIIT).
1995-1998	Bachelor studies in Chemistry at Seethalakshmi Ramasamy College, Affiliated to Bharathidasan University, Trichirapalli
1989-1995	Pre-degree studies at Boiler Plant Girls Higher Secondary School (B.P.G.H.S.S), Trichirapalli

Erklärung

Diese Dissertation wurde selbständig, ohne unerlaubte Hilfe im Sinne von §3 und §5 der Promotionsordnung vom 8. Juli 2002 angefertigt. Bei der Abfassung der Dissertation wurden keine anderen als die darin angegebenen Hilfsmittel benützt.

Zürich, 22. August 2005

V.Sathya Devi